

Research Article

True Orthoimage Generation Using Airborne LiDAR Data with Generative Adversarial Network-Based Deep Learning Model

Young Ha Shin  and Dong-Cheon Lee 

Department of Environment, Energy & Geoinformatics, Sejong University, Seoul 05006, Republic of Korea

Correspondence should be addressed to Dong-Cheon Lee; dcllee@sejong.ac.kr

Received 22 April 2021; Accepted 20 May 2021; Published 14 June 2021

Academic Editor: Hyung-Sup Jung

Copyright © 2021 Young Ha Shin and Dong-Cheon Lee. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Orthoimage, which is geometrically equivalent to a map, is one of the important geospatial products. Displacement and occlusion in optical images are caused by perspective projection, camera tilt, and object relief. A digital surface model (DSM) is essential data for generating true orthoimages to correct displacement and to recover occlusion areas. Light detection and ranging (LiDAR) data collected from an airborne laser scanner (ALS) system is a major source of DSM. The traditional methods require sophisticated procedures to produce a true orthoimage. Most methods utilize 3D coordinates of the DSM and multiview images with overlapping areas for orthorectifying displacement and detecting and recovering occlusion areas. LiDAR point cloud data provides not only 3D coordinates but also intensity information reflected from object surfaces in the georeferenced orthoprojected space. This paper proposes true orthoimage generation based on a generative adversarial network (GAN) deep learning (DL) with the Pix2Pix model using intensity and DSM of the LiDAR data. The major advantage of using LiDAR data is that the data is occlusion-free true orthoimage in terms of projection geometry except in the case of low image quality. Intensive experiments were performed using the benchmark datasets provided by the International Society for Photogrammetry and Remote Sensing (ISPRS). The results demonstrate that the proposed approach could have the capability of efficiently generating true orthoimages directly from LiDAR data. However, it is crucial to find appropriate preprocessing to improve the quality of the intensity of the LiDAR data to produce a higher quality of the true orthoimages.

1. Introduction

True orthoimages are vertical views of the Earth's surface, eliminating distortion of the objects and allowing a view of nearly any point on the ground with a uniform scale. Therefore, the true orthoimages are geometrically equivalent to topographic maps that show true geographic locations of the terrain features. The geometric distortions are caused by the relief displacement of the terrain features (i.e., height variation of the terrain and object surfaces) and perspective projection of the optical cameras that ultimately result in the occlusion areas. In particular, because of the occlusion caused by tall buildings, the surrounding objects are not photographed. Therefore, recovery or compensation of the occlusion areas is crucial in true orthoimage generation [1]. It is a challenging task to automate the entire procedure for generating true orthoimages. Most approaches have focused

on detection and recovery of the occlusion areas. During the last decades, numerous studies for generating true orthoimage have been carried out ever since pixel-based differential rectification has been introduced [2]. Traditional photogrammetric methods require aerial triangulation to obtain exterior orientation parameters (i.e., exposure location and rotation angles) of each aerial image and precise 3D object model data such as the digital building model (DBM) with digital terrain model (DTM) to remove geometric distortion and to detect occlusion areas.

The most crucial tasks are identifying and recovering the occluded areas. In particular, in urban areas with dense high-rise buildings, it is quite difficult to recover occluded areas [3, 4]. As shown in Figures 1 and 2, distortions and occlusions occur in various situations, for example, occlusions on the ground surfaces, surrounding buildings, and even rooftops by superstructures. In addition, sloped or

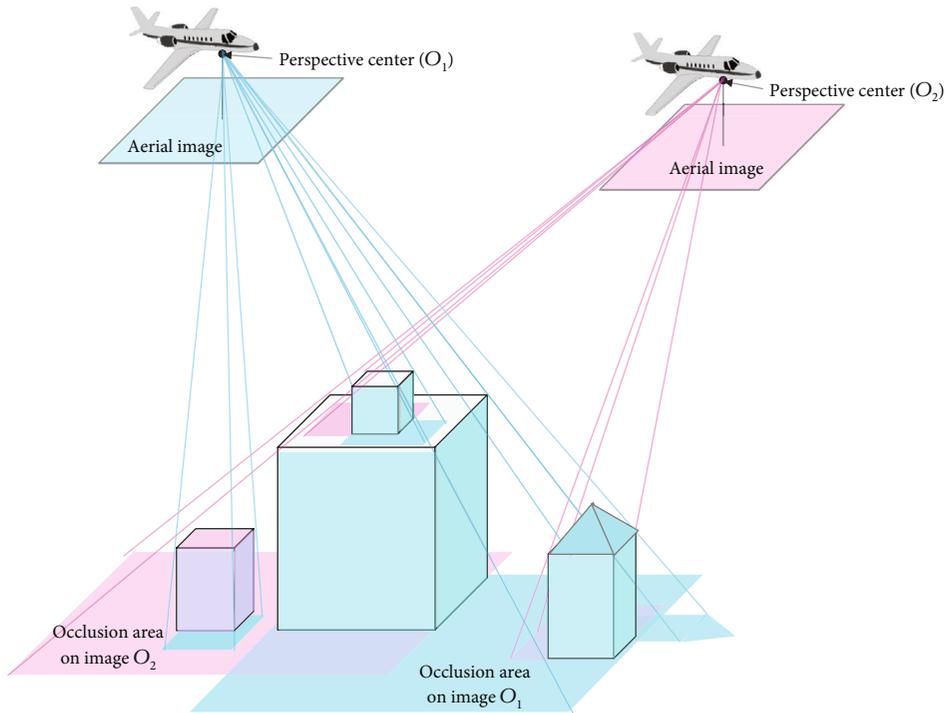


FIGURE 1: Complex nature of occlusion areas on optical aerial images. Different colors represent occlusions of corresponding images. White regions on rooftops depict nonocclusion areas on both images.

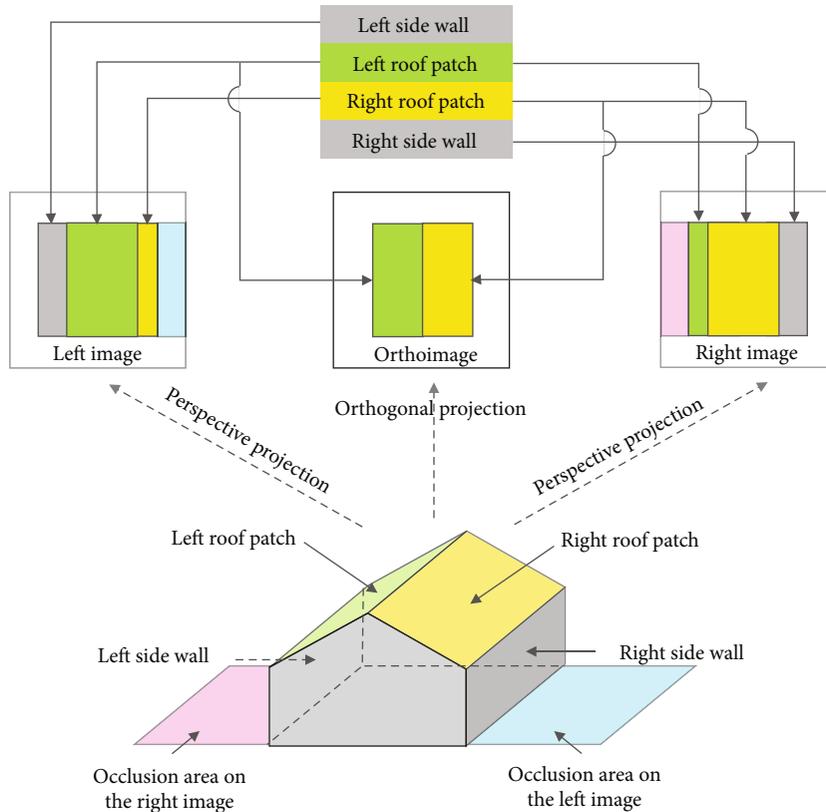


FIGURE 2: Distortion and occlusion in optical images due to relief displacement with central projection.

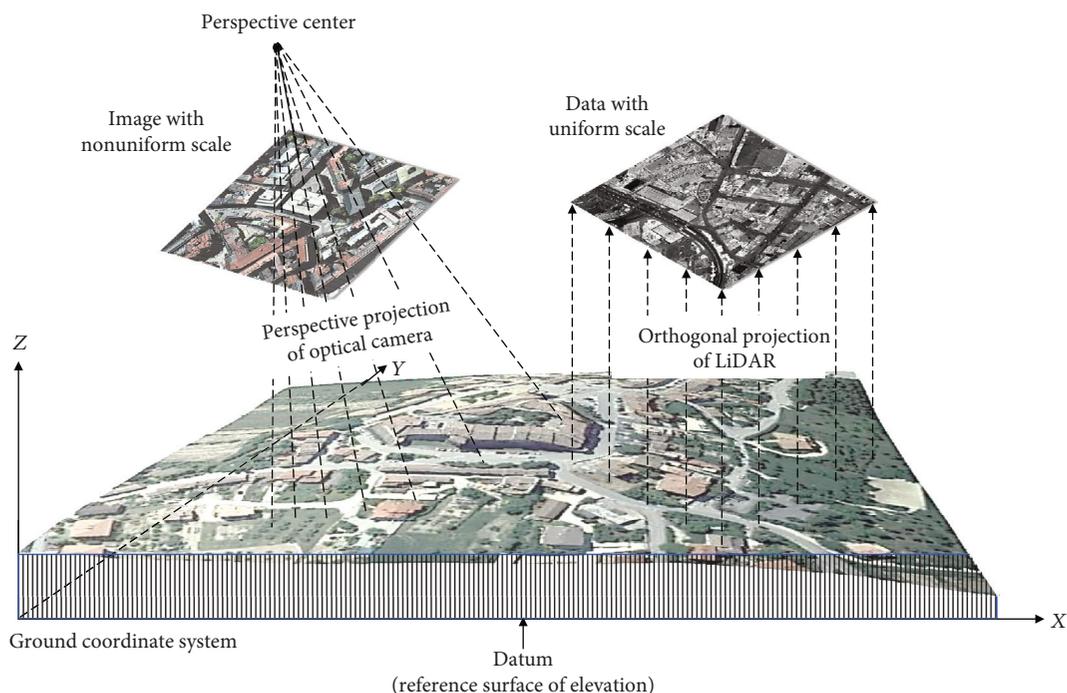


FIGURE 3: Comparison of orthogonal and central projection.

curved surfaces (e.g., gable, pyramid, hip, and dome roof) with varying heights cause various distortions in the images. Considering the correction of all possible details, generating “true” orthoimages automatically involves incredibly difficult processes. In order to identify and recover occluded areas completely, a high level-of-detail (LoD) building model as well as multiview images with accurate EOPs and high overlapping rates is required. Such traditional photogrammetric methods are complicated and costly.

LiDAR data is essentially composed of 3D coordinates and intensity information with a georeferenced orthogonal coordinate system. On the other hand, images are taken with perspective (or central) projection and need an additional process for georeferencing as shown in Figure 3. Therefore, the LiDAR data has a significant advantage in generating true orthoimage because it is not necessary to correct relief displacement and recover occlusions.

Most of the true orthoimage generation methods focus on how to find the areas occluded by the objects and to fill them using overlapped multiview images [5]. Regarding this matter, some of the remarkable approaches are as follows. One of the popular and effective methods is the Z-buffer algorithm for visibility analysis to identify the occlusion areas. The principle of occlusion detection in this method is quite straightforward. The Z-buffer technique identifies occlusions by keeping track of the number of DSM cells projected to a given image pixel. In this algorithm, the distance of each pixel between the projection center and the object surface is stored in a matrix, and the planimetric location of the pixel is recorded in another matrix. These two matrices are defined in an image that represents a visibility map [6, 7]. The Z-buffer method can be applied to the satellite images

obtained from the line scan sensor by searching scan lines corresponding to the equivalent perspective centers [8]. Drawbacks of this technique are sensitivity to the image resolution and DSM cell size. Therefore, false occlusion or false visibility is generated when the image and DSM resolutions are not compatible [3, 9].

Kuzmin et al. [10] proposed a polygon-based method for detecting occlusion areas to generate true orthoimages. In this approach, the differential rectification is applied; then, occlusion areas are detected through the use of polygon surfaces from digital building models (DBMs). Habib et al. [9] and Gharibi and Habib [11] proposed an angle-based occlusion detection method. This method is aimed at improving the accuracy of the occlusion detection using angle and building height using LiDAR data and high-resolution aerial images. They introduced an adaptive radial sweep and spiral sweep approach in which the presence of occlusions can be discerned by sequential checking the off-nadir angles to the direction of light connecting the perspective center to the DSM cells along a radial direction. Yoo and Lee [12] developed a method of patch-based true orthoimage generation while most existing methods have applied a pixel-based approach. The concept of this method is similar to the polygon-based approach in detecting occlusion areas [10]. This method allows mutual recovery of all possible occlusion areas, even including occlusions caused by the superstructures on the rooftops, using multiple images and high level-of-detail digital building model data (i.e., DBM). All images involved in detecting and recovering occlusions can eventually become true orthoimages through compensating each other's occlusions. As a result, multiple true orthoimages are generated for the same scene. Besides the papers

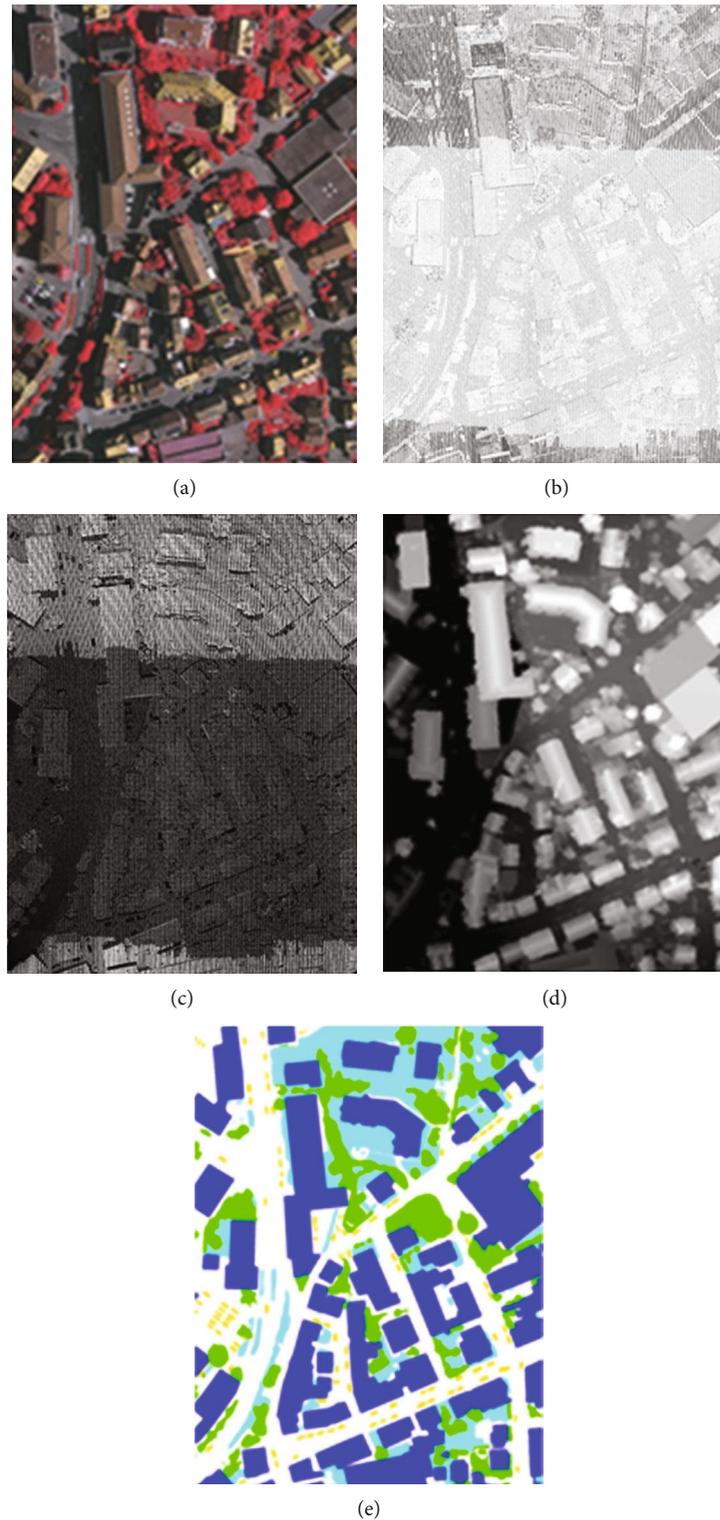


FIGURE 4: Example of Vaihingen datasets: (a) IR true orthoimage; (b) LiDAR intensity; (c) LiDAR DSM; (d) image matching DSM; (e) label data.

mentioned above, there are numerous numbers of literatures concerning true orthoimage generation approaches [13–19]. Generating true orthoimages is not trivial and requires complicated procedures. In consequence, a standard method is currently not available.

This paper is aimed at investigating the feasibility of utilizing DL to generate true orthoimage rather than traditional methods. Only a few studies related to the neural network or DL-based orthoimage generation are available so far. Bagheri and Sadeghian [20] briefly discussed the ability of the genetic

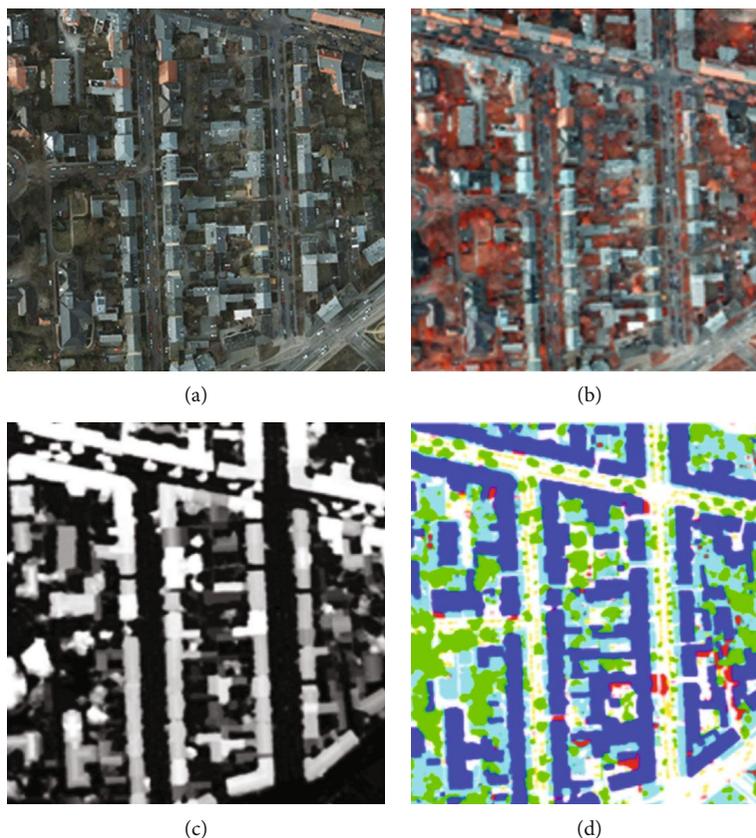


FIGURE 5: Example of Potsdam datasets: (a) natural color (RGB) true orthoimage; (b) IR true orthoimage; (c) image matching DSM; (d) label data.

algorithm and artificial neural network (ANN) for geospatial products such as geometric modeling of satellite images for orthoimage generation and interpolation of DTM. This paper is based on the previous preliminary study on DL-based true orthoimage generation [21]. In this paper, we have carried out intensive and systematic experiments to train DL under various conditions to improve the quality of the true orthoimages. We propose a method based on the Pix2Pix for image-to-image translation that is one of the generative adversarial network (GAN) models. GAN consists of two networks, the generator, which tries to produce results similar to real images (i.e., true orthoimages), and the discriminator, which judges fake and real images until the results are satisfied. Such a mutually adversarial mechanism improves the quality of the results [22]. The training, validation, and test datasets include infrared (IR) orthoimages, intensity and DSM derived from LiDAR data, and label data provided by the International Society for Photogrammetry and Remote Sensing (ISPRS). Finally, generated IR true orthoimages were converted to the pseudonatural color images using a neural network model with prior initial weights instead of random values. The initial weights were adjusted during the training of the model. Generating true orthoimages with DL that could circumvent the complex steps involved in the traditional method would be an attractive

approach even though more improvements in various aspects would be necessary.

2. Datasets and Proposed Method

2.1. Description of Datasets. The Vaihingen and Potsdam datasets provided by ISPRS were used in the experiments [23]. ISPRS datasets are suitable for geoinformatics-related study including this paper since the datasets consist of color infrared (CIR) true orthoimages, airborne LiDAR data, DSM created by image matching, and label data that can be used for training, validation, and test of the DL models (see Figure 4) [24]. The Potsdam datasets consist of IR and natural color (i.e., RGB) true orthoimages, DSM created by image matching, and label data that were not involved in training the DL models but used as the “unseen” data to evaluate the trained DL models. It is noticed that the LiDAR data is not available in the Potsdam datasets (see Figure 5). The label contains object categories for semantic segmentation such as building, road, impervious surface, tree, low vegetation, and water body. The datasets are summarized in Table 1.

The LiDAR data includes 3D coordinates and intensity values of the point clouds. 3D coordinates of the LiDAR data can be used to create DSM. Two types of the DSM (i.e., image matching DSM and LiDAR DSM) were used in the

TABLE 1: Description of datasets.

| Dataset | Image | LiDAR | DSM | Label | GSD (m) | Usage |
|-----------|---|------------------------------|----------------|--------------------|---------|-------------------------------------|
| Vaihingen | IR true orthoimage | Intensity and 3D coordinates | Image matching | Land cover classes | 0.09 | Training, validation, and test data |
| Potsdam | Natural color orthoimage and IR true orthoimage | N/A | Image matching | Land cover classes | 0.05 | Unseen data |

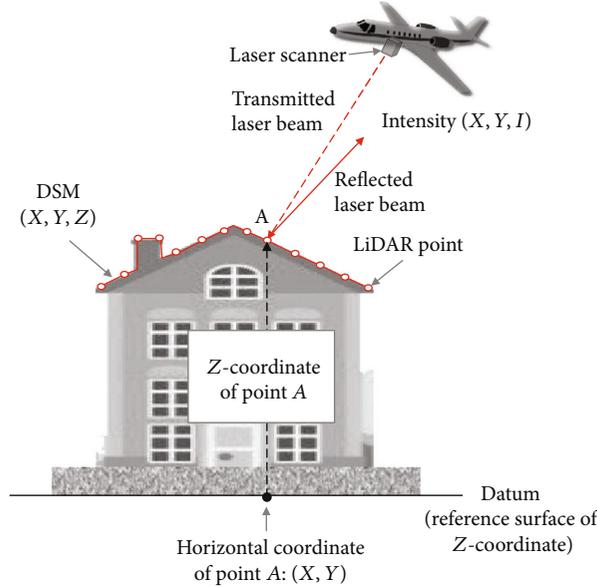


FIGURE 6: Intensity and DSM from LiDAR data.

experiments. The intensity values of the LiDAR data refer to the measure of the return strength of the laser pulses that are reflected from surfaces of the objects. The return intensity is based on the reflectivity of the surface and could be used as supplementary information for feature extraction and classification. DSM and intensity provide 3D geometric and topographic and physical properties, respectively. In these respects, we utilized both intensity (I) and height (Z) of the LiDAR data to train the DL models because intrinsic features of the data play an important role (see Figure 6).

The intensity values can be visualized as black-and-white (B/W) imagery, which is referred to as an intensity map (see Figure 4(b)). However, the quality of the intensity map is poor in terms of spatial and radiometric resolution compared to the optical imagery. If the quality of the intensity maps is high and color-coded, the intensity map itself would be the true orthoimage. These facts gave us the insight to generate the true orthoimages using the LiDAR data using an image-to-image translation DL model.

2.2. LiDAR Dataset Resampling. Since each data has a different ground sample distance (GSD), further processing to make the same GSD of the data is required to train the convolutional neural network- (CNN-) based DL models. Irregularly distributed LiDAR point clouds were resampled to produce regularly gridded DSM. We used inverse distance weighted (IDW) interpolation for the resampling. Figure 7

shows the original LiDAR data of irregularly distributed point clouds, along with resampled data to the regular grid by the interpolation.

Since the point density of the Vaihingen LiDAR data is 4 pts/m^2 , which is equivalent to 0.50 m GSD, the interpolation applied was inconsistent with the resolution of the other datasets (i.e., orthoimages, image matching DSM, and label data) that is 0.09 m . Overlapping is essential in collecting LiDAR data to avoid gaps between strips. It is evident that there are areas of overlap between the adjacent LiDAR strips as shown in Figures 7(a) and 7(c). The apparent visible stripe patterns in the original LiDAR data were removed after resampling.

2.3. Training Dataset Partitioning. Each data was partitioned into tiles by overlapping along both X and Y directions. We created different tile sizes (256×256 , 512×512 , and 1024×1024 pixels) having different overlapping rates (50%, 60%, 70%, and 80%) for the various experiments (see Figure 8). Partitioning with overlapping provides substantial benefits as shown in Figure 9: (1) increasing the number of datasets (i.e., data augmentation) because the DL requires a numerous number of data and (2) improving performance of the DL models by training the same object (or region) under the various situations [25].

In many cases, the label data is used as the ground truth of the categorized land cover features for classification or

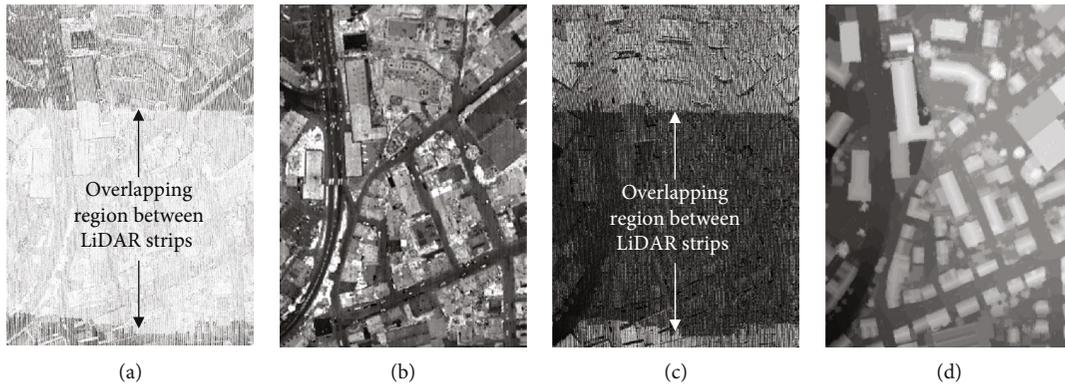


FIGURE 7: Resampling LiDAR data using IDW interpolation for Vaihingen datasets: (a) original LiDAR intensity; (b) resampled LiDAR intensity; (c) original LiDAR 3D coordinates; (d) resampled LiDAR DSM.

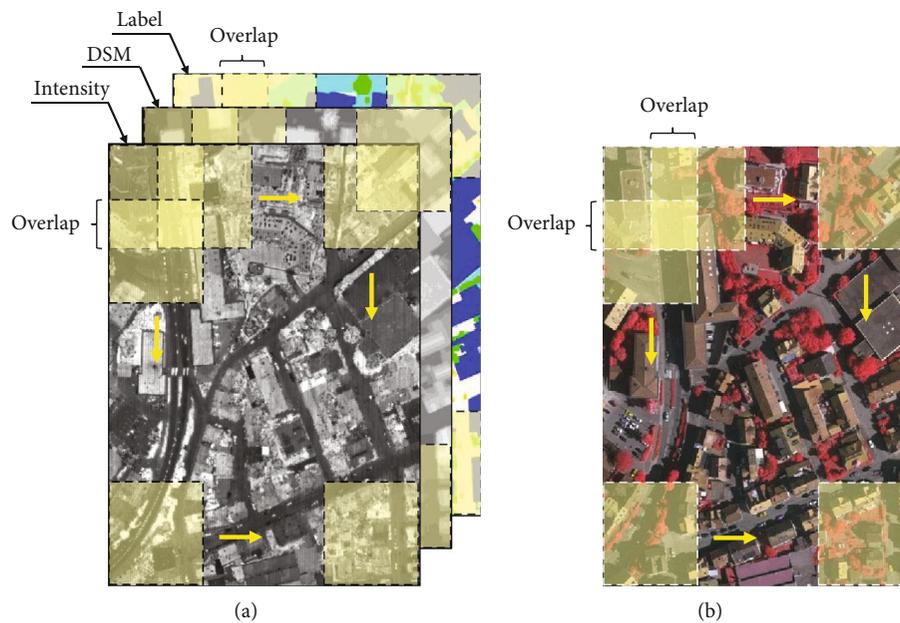


FIGURE 8: Partition of datasets with overlap for both length and width directions: (a) training datasets; (b) ground truth.

semantic segmentation. On the other hand, in this paper, the label data were used as the training data along with the LiDAR datasets while the orthoimages were used as ground truth since we aim to generate orthoimages by training LiDAR and label data using a GAN-based DL model.

2.4. Pix2Pix Model. GANs are algorithmic architectures that are composed of two neural networks, pitting one against the other to generate new synthetic instances of data that can pass for real data. One neural network that is the generator produces new data instances while the other that is the discriminator evaluates them for authenticity, i.e., the discriminator decides whether each instance of data that it reviews belongs to the actual training dataset or not [22]. The GAN models have been frequently used for image style transfer through the mapping from an input image to an output image. In particular, the Pix2Pix as a conditional GAN (cGAN) is well suited for translating an input image into an

output image (i.e., image-to-image translation) [26]. The Pix2Pix is effective for various tasks such as semantic segmentation, generation of maps from aerial images, and colorization of B/W images [27]. As with most cGAN, Pix2Pix also requires pairs of images for training models (i.e., one for input and corresponding ground truth). The challenge of the Pix2Pix is to create paired training datasets while the CycleGANs utilize the unpaired datasets. However, if the distribution of the unpaired datasets of the CycleGANs is not stable, appropriate results might not be created. To improve the performance of the image-to-image transform, U-Net is used in the Pix2Pix. The generator is based on the U-Net architecture that has skip connection between each center of the symmetric layers [28]. Each skipped connection concatenates channels from layer i with channels from layer $n - i$, with n the total number of layers in the network.

In addition, Pix2Pix uses PatchGAN in the discriminator that works on small patches of the input image. Then,

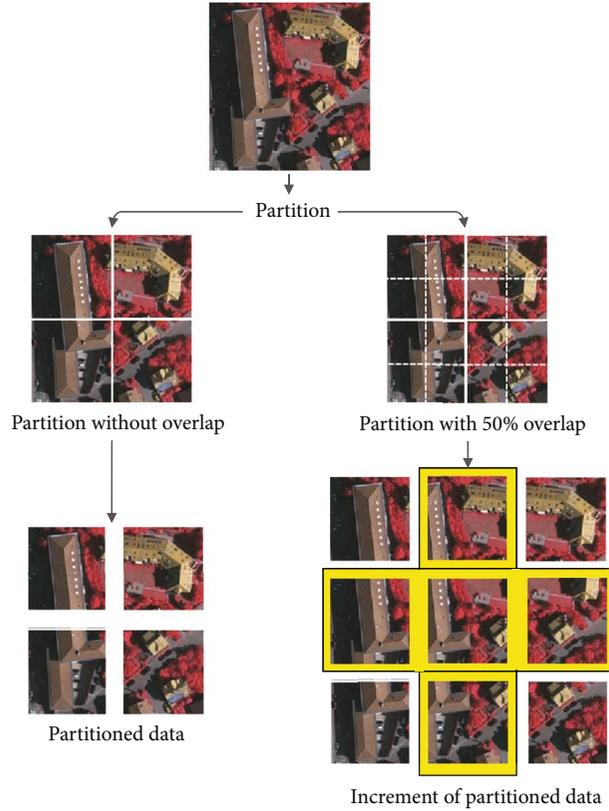


FIGURE 9: Demonstration of dataset increment by partitioning with 50% overlap.

averaging is performed to decide whether the entire image is real or fake. The advantage of using PatchGAN is that the discriminator captures high frequencies in the image by straining the model to focus on the local image patches. Therefore, the convolutional layer will only be receptive to the pattern at the patch size [26, 29]. Both generator and discriminator use modules of the form “convolution-batch normalization-activation function.” The number of the convolution layers in the generator depends on the size of the input images because the convolutions are performed up to 1×1 pixel image. The image size is reduced by half since stride is 2 for pooling after convolution. For an example of a 512×512 image (i.e., $2^9 \times 2^9$ pixels), nine convolution layers are required. Figure 10 shows the Pix2Pix model and workflow of the proposed method. Figures 11 and 12 show the generator and discriminator of the Pix2Pix model, respectively.

ReLU is the most commonly used activation function in CNNs. Specifically, Pix2Pix utilizes the plane ReLU (see Figure 13(a) and Equation (1)) in the encoder and the leaky ReLU with slope 0.2 (see Figure 13(b) and Equation (2)) in the decoder of the generator, while leaky ReLUs also with slope 0.2 are used in the discriminator. It is known that using leaky ReLU is beneficial because of fixing the dying ReLU problem and speeding up the training [30, 31]:

$$\text{ReLU} : \phi(x) = \begin{cases} 0, & \text{for } x \leq 0, \\ x, & \text{for } x > 0, \end{cases} \quad (1)$$

$$\text{Leaky ReLU} : \phi(x) = \begin{cases} ax, & \text{for } x \leq 0, \\ x, & \text{for } x > 0, \end{cases} \quad (2)$$

where a is usually a small constant value (e.g., $a = 0.2$).

2.5. Pix2Pix Model Training Strategies. The Vaihingen datasets were used for training, validation, and testing of the DL models. The Potsdam datasets were used as unseen datasets to evaluate the models trained with the Vaihingen datasets. In order to analyze the performance of the DL models and eventually to find the optimal case, we carried out intensive experiments under various conditions in terms of the data type, data combination, partitioned tile size, and overlap rate. As for the label, each color of the original label images represents a specific land cover feature such as blue for buildings, green for trees, cyan for low vegetation, yellow for cars, and red for water bodies. The original color label images as well as converted to B/W label images were used for training to compare results from each label image (see Figure 14).

The training schemes were categorized based on the following cases (see Figure 15).

- (1) Individual datasets: [intensity], [DSM], [color label], and [B/W label]
- (2) Combined datasets: [intensity+DSM], [intensity+B/W label], [DSM+B/W label], and [intensity+DSM+B/W label]

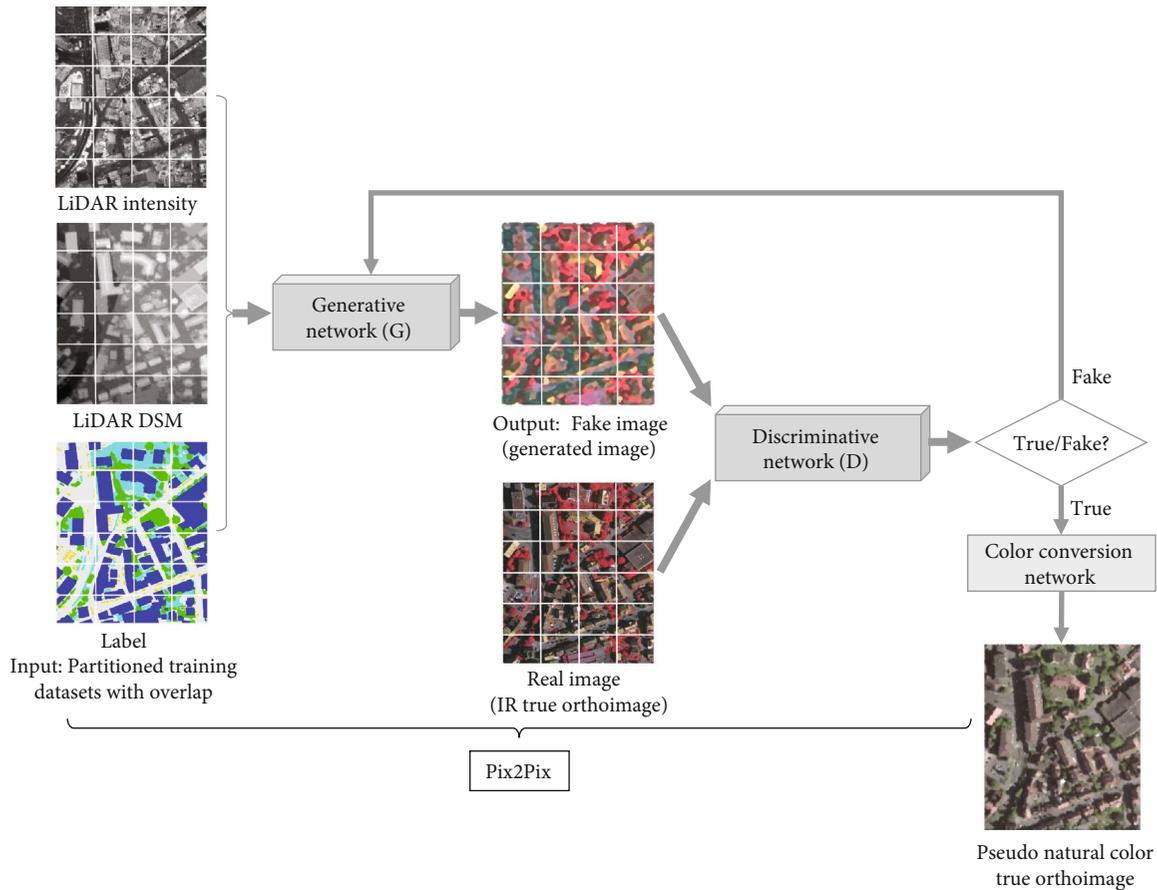


FIGURE 10: Workflow of the proposed method for true orthoimage generation using the DL model.

(3) Tile size: 256×256 , 512×512 , and 1024×1024

(4) Overlap rate: 50%, 60%, 70%, and 80%

Analyzing a variety of the results through the intensive experiments under the various conditions might be helpful to determine the optimal case. The test data and unseen data from the Vaihingen and Potsdam datasets, respectively, were inputted into the trained models. The optimal trained models would be used to the datasets from other areas (see Figure 16).

2.6. Infrared Image to Natural Color Image Conversion. A natural color (i.e., normal color) composite is an image displaying a combination of the primary visible red (R), green (G), and blue (B) components (or bands) to the corresponding channels that resemble what humans observe. On the other hand, an IR image is composed of the near-IR (NIR), R, and G bands. Since the IR images are false color composites, it is often necessary to convert the IR images to natural color images, so-called pseudonatural color images [32]. We performed the spectral conversion from IR to RGB color image using a simple neural network. Each IR true orthoimage from the test data of the Vaihingen area and unseen data of the Potsdam area was converted to the natural color

images using a neural network model (see Figure 17) based on the spectral conversion equations:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} w_{11} & w_{21} & w_{31} \\ w_{12} & w_{22} & w_{32} \\ w_{13} & w_{23} & w_{33} \end{bmatrix} \begin{bmatrix} \text{NIR} \\ R \\ G \end{bmatrix}, \quad (3)$$

where $[R, G, B]^T$ is output bands of a pseudonatural color image and $[\text{NIR}, R, G]^T$ is input bands of an IR image. w_{ij} are weights of the spectral conversion matrix to be determined by training the neural network.

The target images (i.e., natural color images) for the color conversion network training were taken from the aerial imagery of Microsoft Bing Maps. In general, training of the neural network model starts with random initial weights. However, performance of the training would be improved if reasonable initial weights are available. We adopted the pseudo color conversion coefficients used in Satellite Pour l'Observation de la Terre (SPOT) multispectral satellite imagery as initial weights that are $w_{11} = 0.1$, $w_{21} = 0.9$, $w_{31} = 0.0$, $w_{12} = 0.3$, $w_{22} = 0.0$, $w_{32} = 0.7$, $w_{13} = 0.1$, $w_{23} = 0.1$, and $w_{33} = 0.8$. The sum of the weights for each raw should be

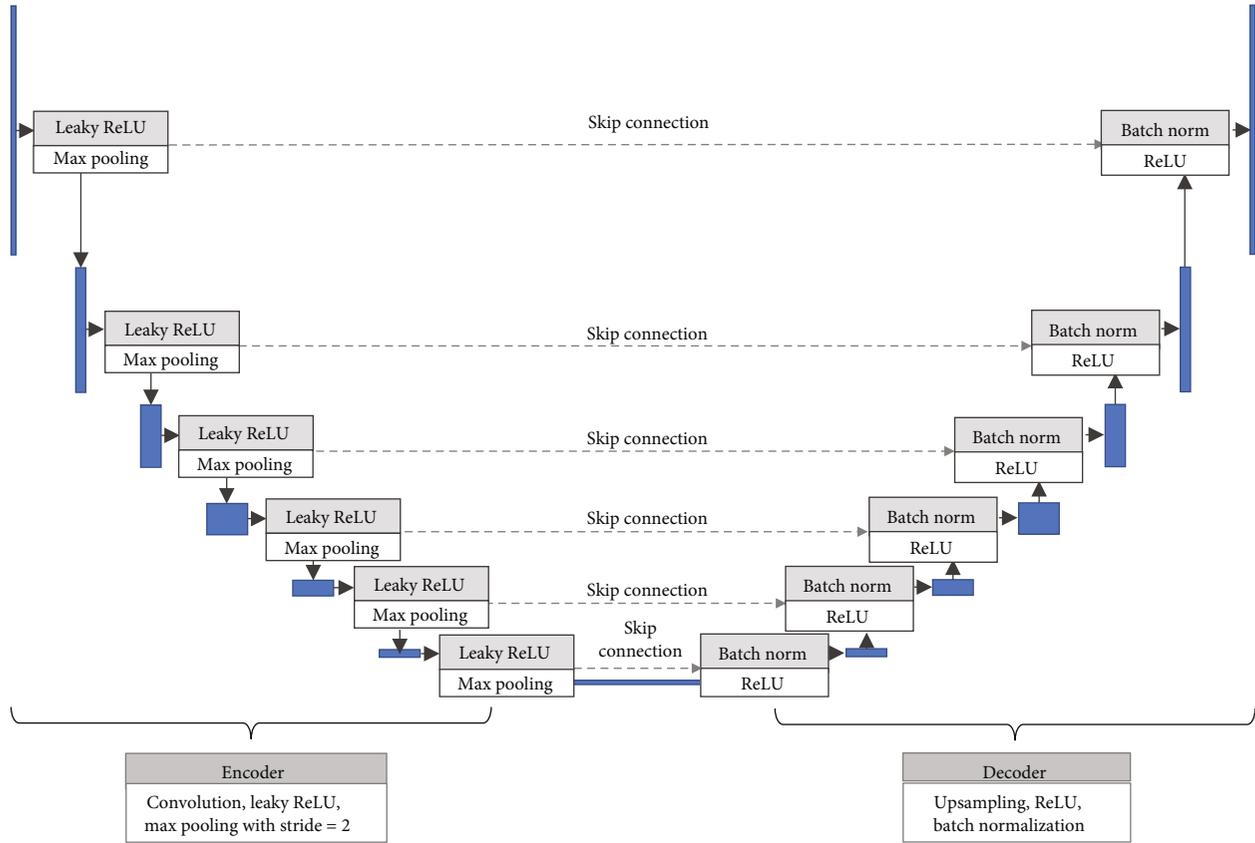


FIGURE 11: Generative network of the Pix2Pix model.

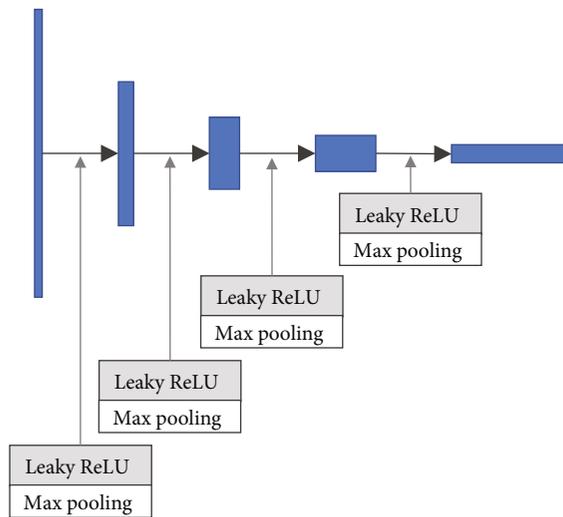


FIGURE 12: Discriminative network of the Pix2Pix model.

1 because the brightness of the images should not change. The weights were updated through the backpropagation.

3. Results and Analysis

As described in the training strategy, experiments using different types of training datasets under various conditions

were conducted to generate true orthoimages and to analyze performance of the DL models. The performance was evaluated with plots of the loss for epoch and the Fréchet inception distance (FID) and structural similarity index measure (SSIM) that have been frequently used as evaluation measures of the GAN-based models [33–36]. FID is a metric to assess the quality of the images created by the generator. It

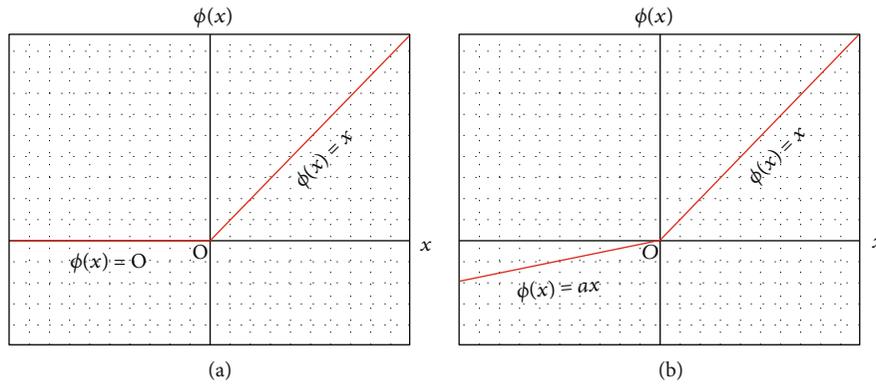


FIGURE 13: Activation functions used in Pix2Pix: (a) ReLU; (b) leaky ReLU.

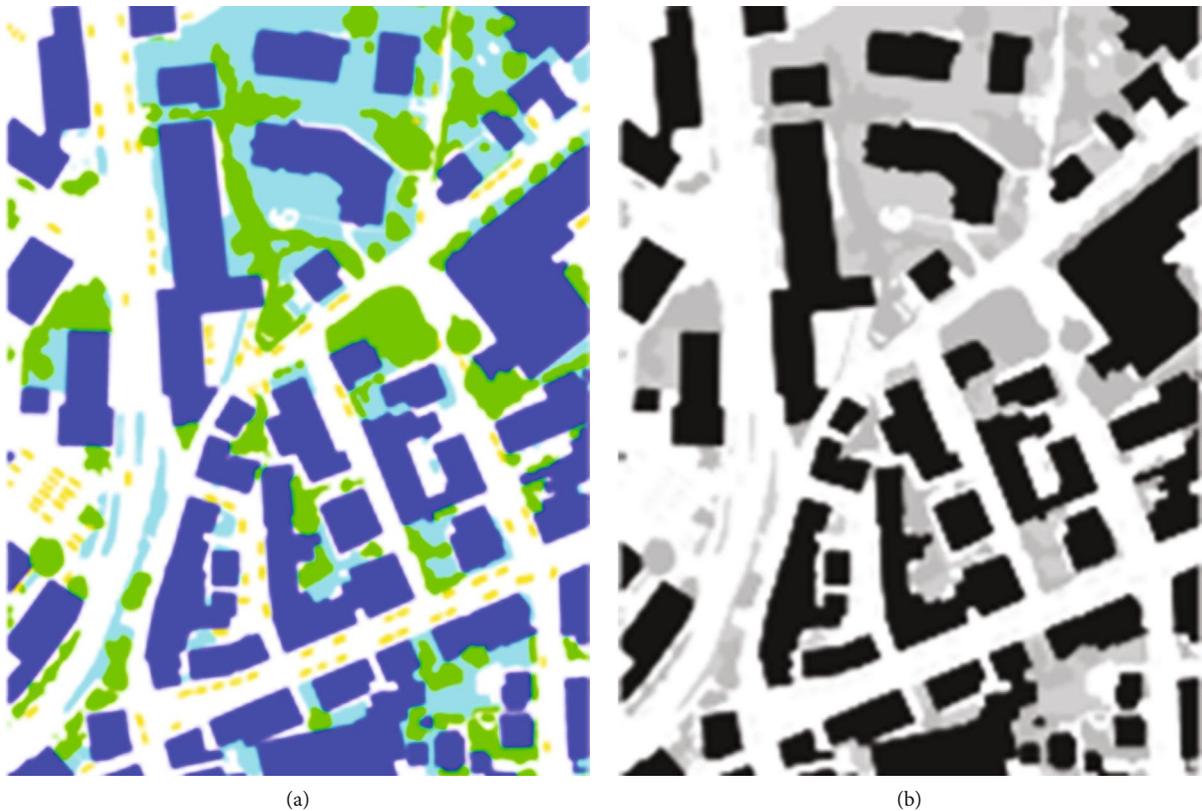


FIGURE 14: Label datasets: (a) color label; (b) B/W label.

measures the distance between feature vectors calculated for the generated images (i.e., fake images) and real images (i.e., ground truth) based on the mean and covariance. A lower FID indicates better quality of the generated image. The SSIM is a perception-based method and has been used for measuring the similarity between two images based on an initial uncompressed or distortion-free image as reference. The SSIM is also calculated using the mean and covariance as FID. The maximum value of the SSIM is 1, which indicates that the two images are perfectly structurally similar, while a value of 0 indicates no structural similarity. However, analyzing or assessing the quality of images based on the statisti-

cal or quantitative metric might be insufficient since such metrics represent overall quality without local details. Therefore, we carried out visual inspection for each case.

Training of the Pix2Pix models was performed with a maximum of 200 epochs. The learning rates were set to 0.0002 for the 1st to 100th epoch and then linearly decreased up to 0.000002 after the 101st to 200th epoch. The loss graphs for LiDAR intensity, LiDAR DSM, and label image are shown in Figure 18. The loss for each data rapidly decreased in the early epochs. In our experiments, the overall losses showed a similar pattern, i.e., the losses were getting stable and converged after the 60th epoch.

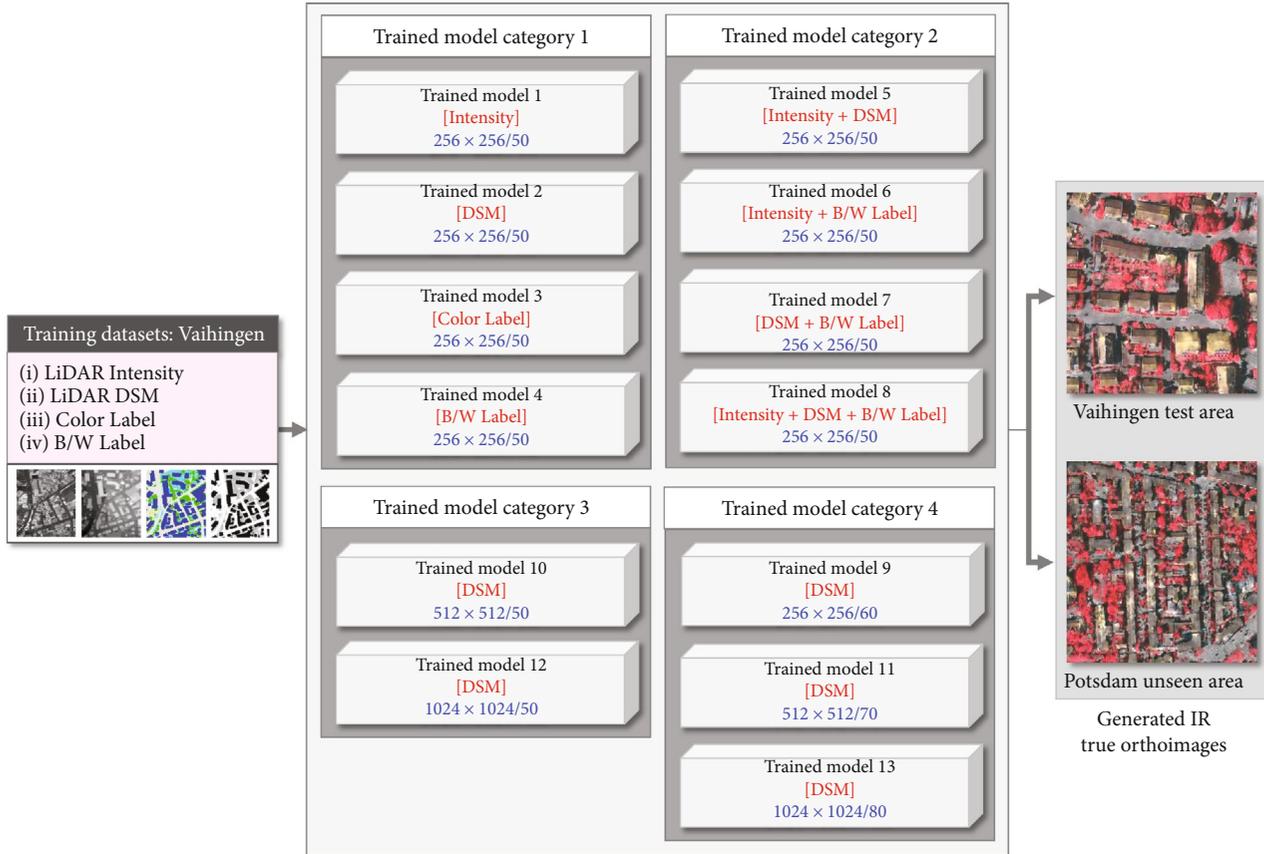


FIGURE 15: True orthoimage generation for test and unseen datasets using trained models. The numbers in each trained model represent tile size and overlap rate.

The loss graph in terms of the root mean squared error (MSE) for the IR to natural color conversion is presented in Figure 19.

The main purpose of the experiments is to determine the optimal training condition by analyzing the influences of the data characteristics, tile size and overlap of the data partitioning, and combinations of the different types of data for the training. The results and analyses of the generated orthoimages and pseudonatural color images for the test and unseen datasets are presented in various aspects.

3.1. Results of True Orthoimage Generation for Test Datasets.

The first series of experiments is for examining the results from training each data independently with fixed partitioned tile size (256×256) and the overlap rate (50%) (see Figures 20–23). Since the LiDAR intensity has low radiometric resolution and noisy stripe pattern, the result is not satisfactory. Compared to the real true orthoimage (see Figure 20(a)), it is difficult to distinguish between buildings and trees (see Figure 20(b)). The result of the LiDAR DSM is better than the result of the intensity because DSM contains geometric information such as shape and surface slope of the objects (see Figure 20(c)). The label is an image of classified land cover features, and each class is represented with an arbitrary color. The experiments were conducted with color label (see Figure 21(b)) and B/W label (see

Figure 21(c)) to investigate the effect of the label color scheme. The results show that there is no significant difference between color and B/W labels (see Experiments 1 to 4 in Table 2). The label has different characteristics from the LiDAR intensity and DSM. The advantage of using the label for training DL model is to produce clear and sharp boundaries of the buildings (see Figures 21(b) and 21(c)) while LiDAR data does not produce clear boundaries due to resampling (see Figures 20(b) and 20(c)).

The second series of experiments is for exploring the effect of combining individual data. We carried out experiments with all possible combinations of the datasets. Combining LiDAR intensity and LiDAR DSM (see Figure 22(b)) shows improved results compared with LiDAR intensity only (see Figure 20(b)). Combining LiDAR intensity and B/W label (see Figure 22(c)) provided better results than combining LiDAR intensity and LiDAR DSM. Moreover, combining LiDAR DSM and B/W label (see Figure 23(b)) resulted in improved true orthoimages. In this case, the shape of the roofs is identifiable, particularly the gable roofs.

Finally, all three data (i.e., LiDAR intensity, LiDAR DSM, and B/W label) were combined (see Figure 23(c)). The result was improved compared to combining LiDAR DSM and B/W label (see Figure 23(b)). Therefore, combining all datasets brought synergistic effect to improve the performance of the training. This fact is also shown in Table 2. Experiment 8,

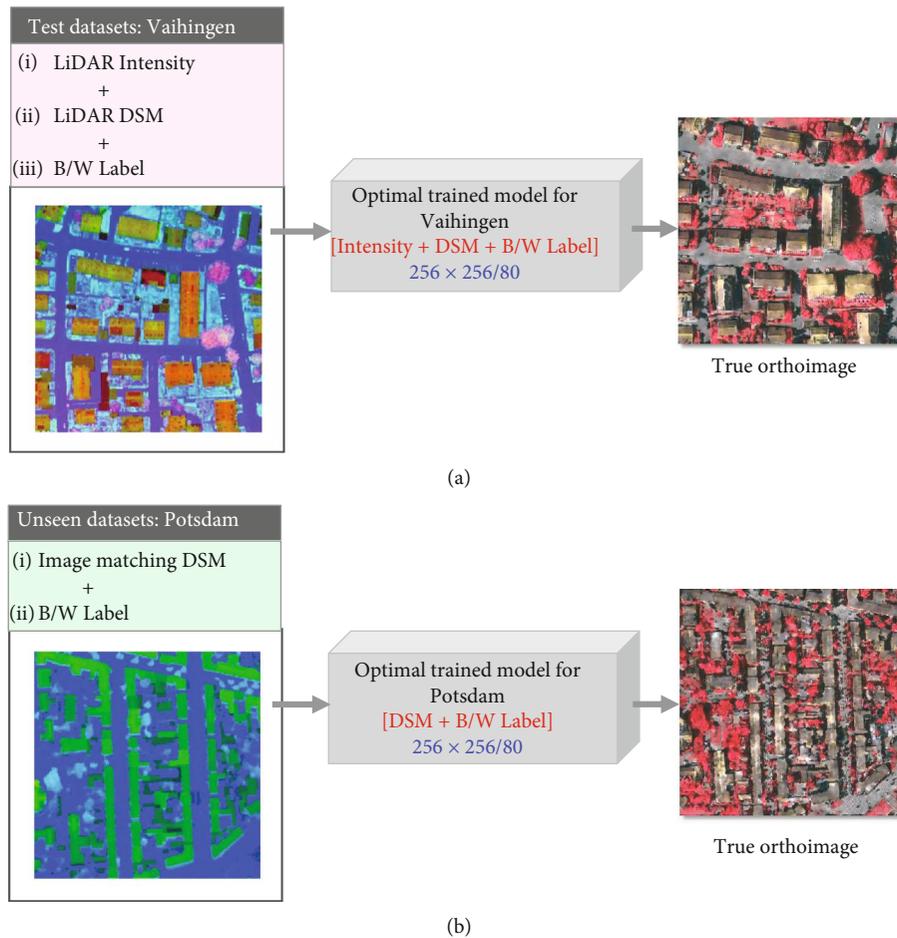


FIGURE 16: True orthoimages from optimal trained models: (a) Vaihingen test datasets; (b) Potsdam unseen datasets.

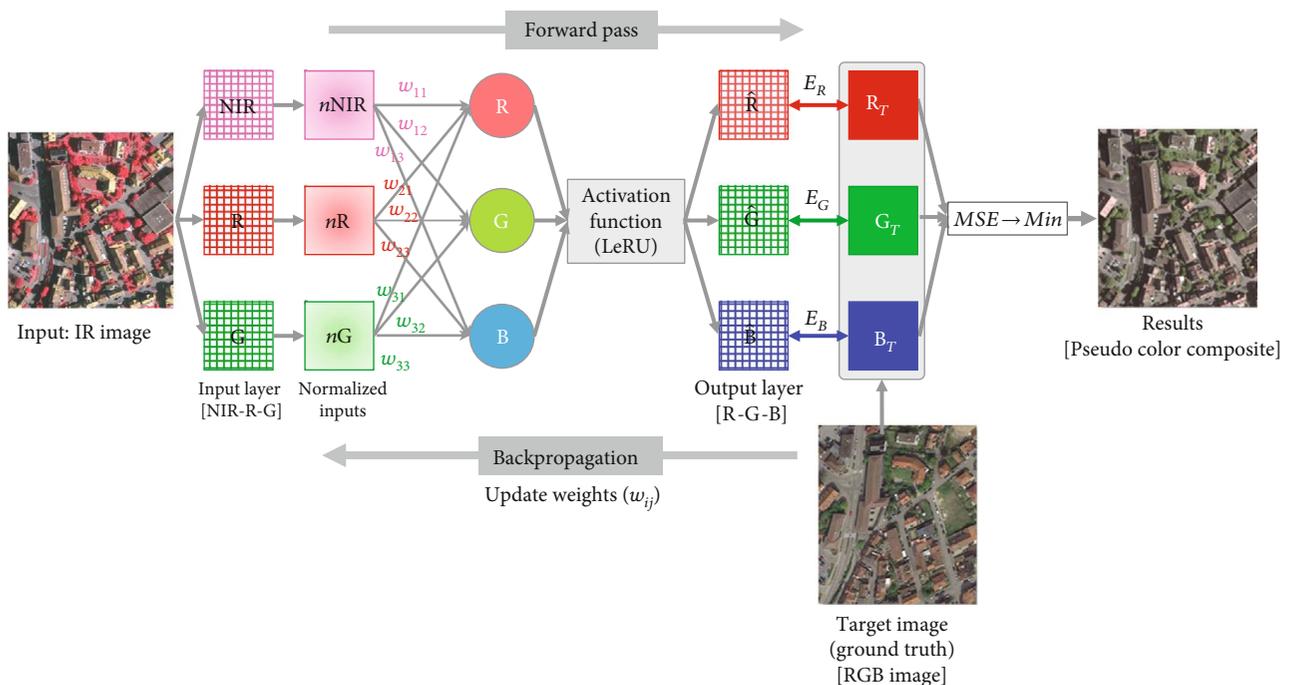


FIGURE 17: A neural network for IR to pseudonatural color image conversion.

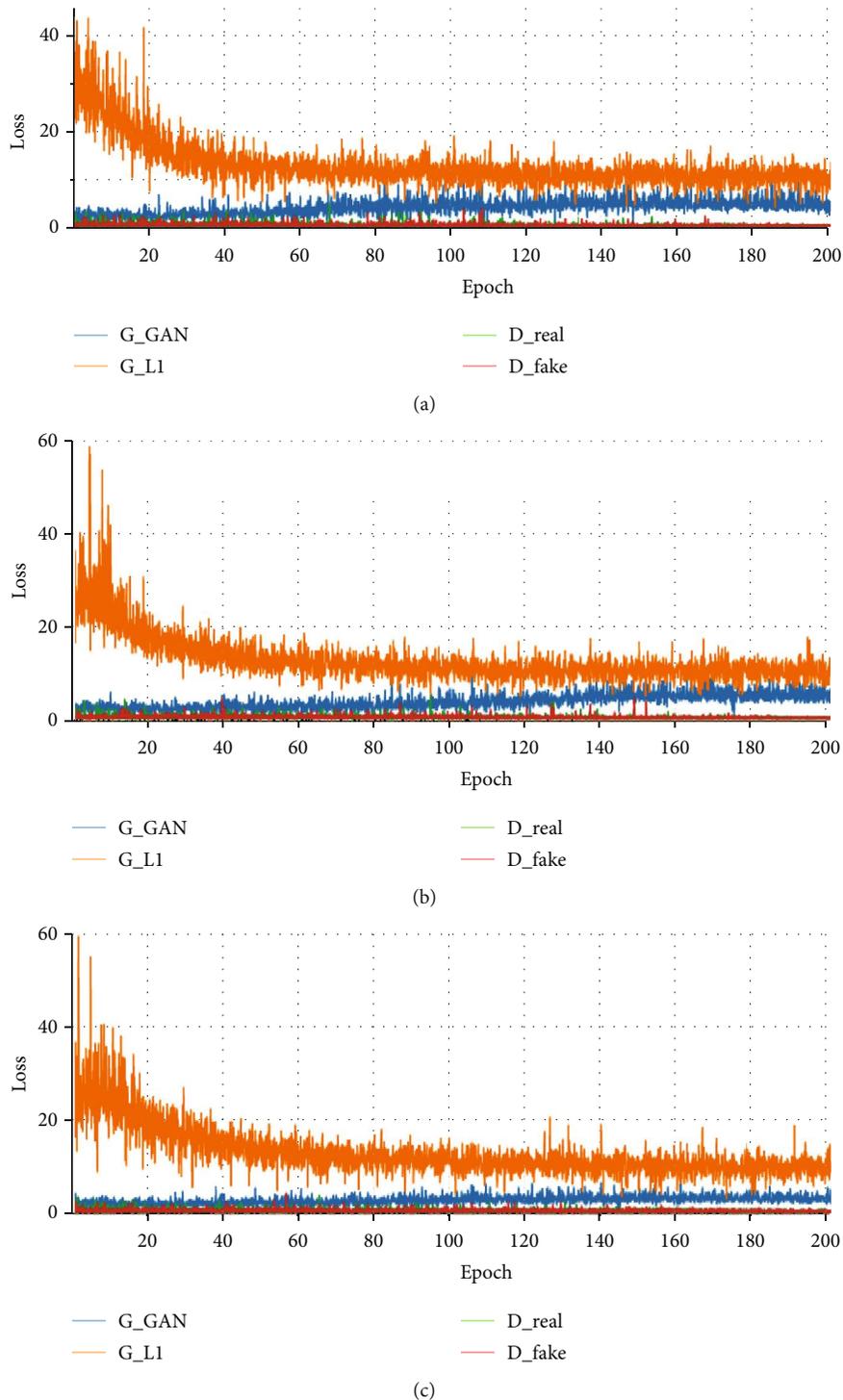


FIGURE 18: Loss graphs of Pix2Pix model training: (a) LiDAR intensity; (b) LiDAR DSM; (c) label.

which is combining all datasets, provides the smallest FID (i.e., 131.83) and the largest SSM (i.e., 0.0434) among the individual dataset training and the combining training cases (see Experiments 1 to 8 in Table 2).

The third series of experiments is for comparing the influence of the overlap in data partitioning. This experiment was carried out using LiDAR DSM. The results were analyzed

for 50% and 60% overlap with a fixed tile size of 256×256 . For the visual comparison of the results, the real true ortho-image is provided in Figure 24(a). The larger overlap results in dataset augmentation that is required for training DL models. 60% overlap (see Figure 24(c)) provides better results than 50% overlap (see Figure 24(b)). The roof shapes are improved by increasing overlap. Experiment 2 (i.e., FID =

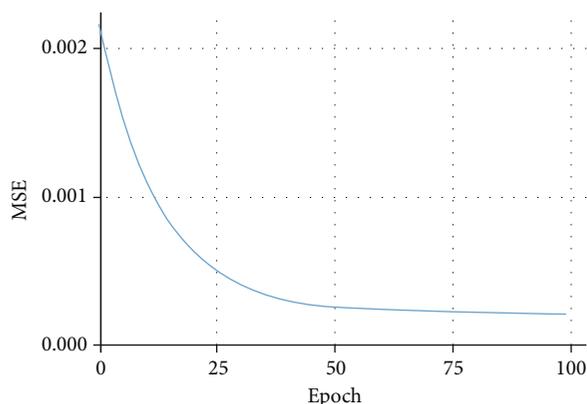


FIGURE 19: Loss graph of the color conversion network model.

231.74 and SSIM = 0.401) and Experiment 9 (i.e., FID = 197.63 and SSIM = 0.427) in Table 2 show that the larger overlap provides better results. More experiments for the tile size of 512×512 with 50% and 70% overlaps (see Figures 25(b) and 25(c)) and for the file size of 1024×1024 with 50% and 80% overlaps (see Figures 26(b) and 26(c)) were performed. Similar results were obtained for all cases (see FID and SSIM of Experiments 10 and 11 and Experiments 12 and 13 in Table 2).

The fourth series of experiments are aimed at finding the optimal tile size. The real true orthoimages are shown in Figures 25(a) and 26(a) to compare the results from the experiments under different conditions. The results for different tile sizes are presented in Figures 20(c), 25(b), and 26(b) corresponding to the tile size of 256×256 , 512×512 , and 1024×1024 , respectively. The results look similar to each other even though the tile size of 256×256 might be the best one according to both FID (i.e., 231.74) and SSIM (i.e., 0.401) in Table 2. On the other hand, FID and SSIM for tile size of 512×512 are 316.40 and 0.388, respectively. FID and SSIM for tile size of 1024×1024 are 400.89 and 0.363, respectively.

Based on the experiments considering various aspects including tile size and overlap rate, characteristics of the training datasets, quality evaluation metrics (i.e., FID and SSIM), and visual inspection, using all data (LiDAR intensity, LiDAR DSM, and B/W label) with 256×256 tile size and 80% overlap provided the best result with the smallest FID (i.e., 117.67) and the largest SSIM (i.e., 0.505) (see Figure 27(b) and Experiment 15 in Table 2). The second-best result was obtained by combining DSM and B/W label (see Experiment 14 in Table 2) with FID of 132.35 and SSIM of 0.436, and this trained model was applied to the unseen data of the Potsdam since the intensity data is not available in the Potsdam datasets. Figure 28 represents Table 2 with bar graphs for better analysis and interpretation of the results.

Image quality evaluation considers various elements such as color, tone, brightness, and contrast. Furthermore, evaluation of the true orthoimages is a complicated issue because assessment of the appropriate recovery of the occlusion areas is essential. For this reason, the difference between the real

true and generated true orthoimages might be objective and both quantitative and qualitative ways of evaluating orthoimages. Figure 27 visualizes the difference between a real true orthoimage and a generated true orthoimage of the test datasets. The darker areas in Figure 27(c) represent regions of significant difference between the orthoimages, indicating that the image is incorrectly generated, such as the color and the tone of some buildings. In particular, the incorrectly generated regions are along the building boundaries due to the interpolation for resampling of the LiDAR data. However, there are limitations to evaluate the quality of the generated image based on the difference since the disparity might not reflect all factors of the image quality.

3.2. Results of True Orthoimage Generation for Unseen Datasets. The results of the generated true orthoimages for the Potsdam unseen datasets are presented in Figures 29–34. The intensity data could not be used, and the image matching DSM was used in lieu of the LiDAR DSM since the LiDAR data is not available in Potsdam datasets. The results of the unseen datasets were obtained by reusing the pretrained models that were trained with Vaihingen training datasets. In other words, neither training nor fine tuning of the hyperparameters was performed to generate the true orthoimage. Therefore, the quality of the results from unseen datasets is not expected to be better than the case of the test datasets. Analyses of the results from the unseen datasets might not be consistent with the results from the test datasets because of the respective datasets belonging to different places with different characteristics such as image tone, image scale, topology, and object shape. The series of experiments for the unseen datasets are the same order as the test datasets, i.e., analysis of the results from (1) the use of each data independently, (2) combination of individual data, (3) different overlap rates, and (4) different tile sizes.

The result using DSM (see Figure 29(b)) was visually better than using other data (i.e., color label and B/W label). Also, both FID and SSIM indicate that using the DSM is better than using the other data. The smallest FID (i.e., 451.39) and the largest SSIM (i.e., 0.375) were obtained from DSM (see Experiments 1, 2, and 3 in Table 3). In addition, using the combined dataset of DSM and B/W label did not improve the results (see Figure 30(c) and Experiment 4 in Table 3).

As for the influence of the overlap, we compared the overlap rates between 50% and 60% with tile size of 256×256 (see Figures 31(b) and 31(c) and Experiments 1 and 5 in Table 3) for DSM. Even though it is not easy to tell the difference between the results according to the FID and SSIM, 60% overlap resulted in more or less clear edges of the buildings than 50% overlap by visual inspection. Similar results were obtained for the overlap rates between 50% and 70% with tile size of 512×512 (see Figures 32(b) and 32(c) and Experiments 6 and 7 in Table 2). However, for the case of overlap rates of 50% and 80% with tile size of 1024×1024 , 80% overlap produced better results than 50% overlap (see Figures 33(b) and 33(c) and Experiments 8 and 9 in Table 2).

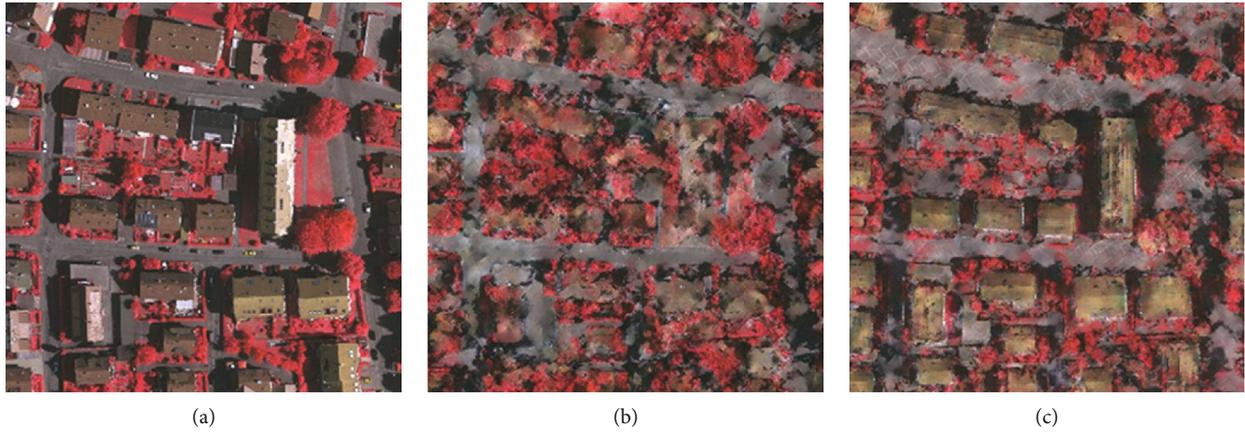


FIGURE 20: Generated true orthoimages with 256×256 tile size and 50% overlap: (a) real true orthoimage; (b) LiDAR intensity; (c) LiDAR DSM.

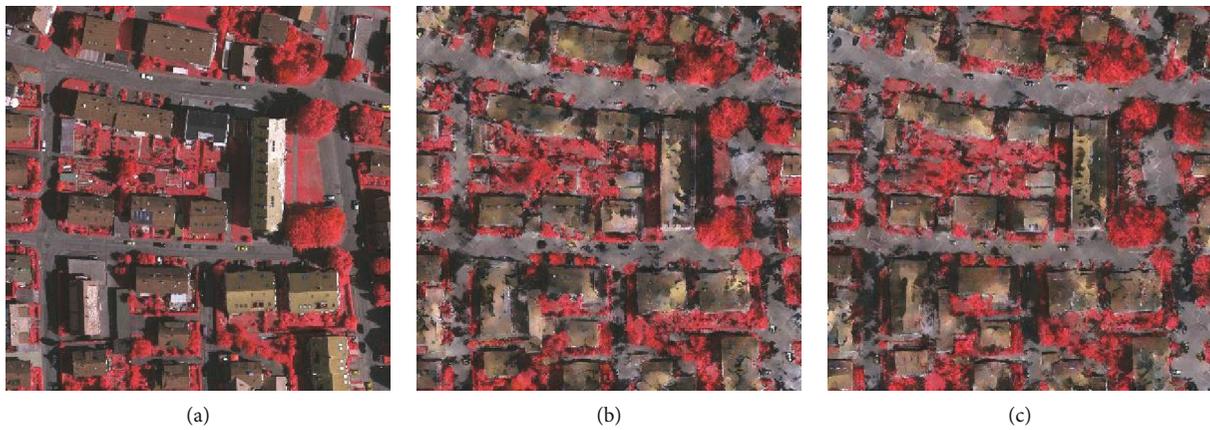


FIGURE 21: Generated true orthoimages with 256×256 tile size and 50% overlap: (a) real true orthoimage; (b) color label; (c) B/W label.

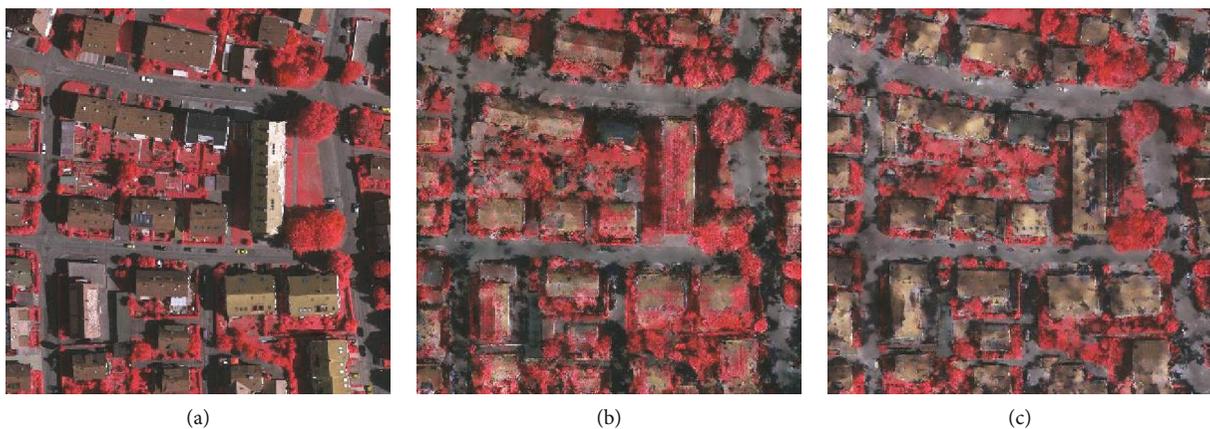


FIGURE 22: Generated true orthoimages with 256×256 tile size and 50% overlap: (a) real true orthoimage; (b) LiDAR intensity and LiDAR DSM; (c) LiDAR intensity and B/W label.

The best result was obtained from the model trained using all of the data (i.e., LiDAR intensity, LiDAR DSM, and B/W label) of the Vaihingen datasets with 256×256 and 80% overlap. Nevertheless, this trained model could

not be applied to the unseen data because Potsdam datasets do not include LiDAR data. Therefore, the model trained using image matching DSM and B/W label, which is the second-best case of the experiments (see Experiment 14 in

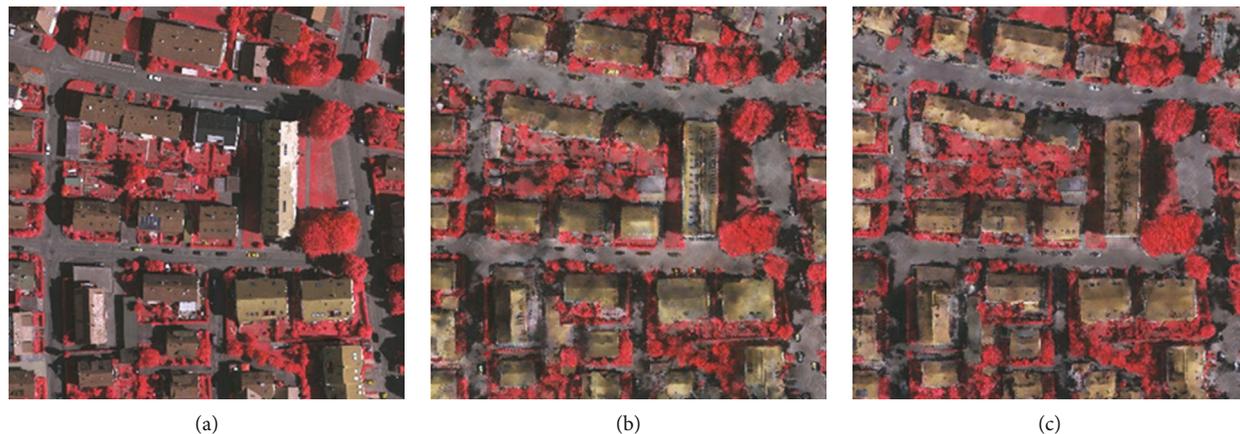


FIGURE 23: Generated true orthoimages with 256×256 tile size and 50% overlap: (a) real true orthoimage; (b) LiDAR DSM and B/W label; (c) LiDAR intensity; LiDAR DSM and B/W label.

TABLE 2: Evaluation of training performance for Vaihingen test datasets.

| Experiment | Training data type | Tile size (pixel) | Overlap (%) | FID | SSIM |
|------------|-------------------------|--------------------|-------------|--------|-------|
| 1 | Intensity | 256×256 | 50 | 353.18 | 0.371 |
| 2 | DSM | 256×256 | 50 | 231.74 | 0.401 |
| 3 | Color label | 256×256 | 50 | 222.51 | 0.379 |
| 4 | B/W label | 256×256 | 50 | 194.76 | 0.383 |
| 5 | Intensity+DSM | 256×256 | 50 | 250.44 | 0.396 |
| 6 | Intensity+B/W label | 256×256 | 50 | 218.01 | 0.421 |
| 7 | DSM+B/W label | 256×256 | 50 | 214.77 | 0.431 |
| 8 | Intensity+DSM+B/W label | 256×256 | 50 | 131.81 | 0.434 |
| 9 | DSM | 256×256 | 60 | 197.63 | 0.427 |
| 10 | DSM | 512×512 | 50 | 258.59 | 0.395 |
| 11 | DSM | 512×512 | 70 | 174.79 | 0.406 |
| 12 | DSM | 1024×1024 | 50 | 369.21 | 0.368 |
| 13 | DSM | 1024×1024 | 80 | 148.41 | 0.377 |
| 14 | DSM+B/W label | 256×256 | 80 | 132.35 | 0.436 |
| 15 | Intensity+DSM+B/W label | 256×256 | 80 | 117.67 | 0.505 |

Table 2), was applied to the unseen data. As portrayed in Table 3, the smallest FID and SSIM values of 373.03 and 0.387, respectively, were produced in Experiment 10 which corresponds to the approach that produced the best outcome. The result is shown in Figure 34(b), and the difference between the real true orthoimage and generated true orthoimage is visualized (see Figure 34(c)). Figure 35 represents Table 3 with bar graphs.

3.3. Results of Pseudocolor Image Generation. Conversion of the IR image to the pseudonatural image depends mostly on the target image used as ground truth such as time (or season), illumination condition, and camera characteristics. Since IR is beyond the visible spectrum, CIR images are displayed as the false color composite. The invisible NIR light of CIR can be visible by shifting the NIR light and the primary

colors over. NIR wavelengths become visible as red while red wavelengths appear as green and green as blue. Blue wavelengths are shifted out of the visible portion of the spectrum, and so they appear as black. Most of the vegetation appears red while water generally appears black with artificial structures like buildings and roads showing as a light blue-green on the CIR images.

We carried out the color conversion using a neural network by training natural color images. The results of the test and unseen images are shown in Figures 36 and 37, respectively. The initial and final weights after training the neural networks are presented in Table 4. Regardless, the results of the pseudocolor images might not be the same as the natural color image of the primary visible red, green, and blue. In addition, various color schemes could be utilized by applying different weight constraints of the neural network model.

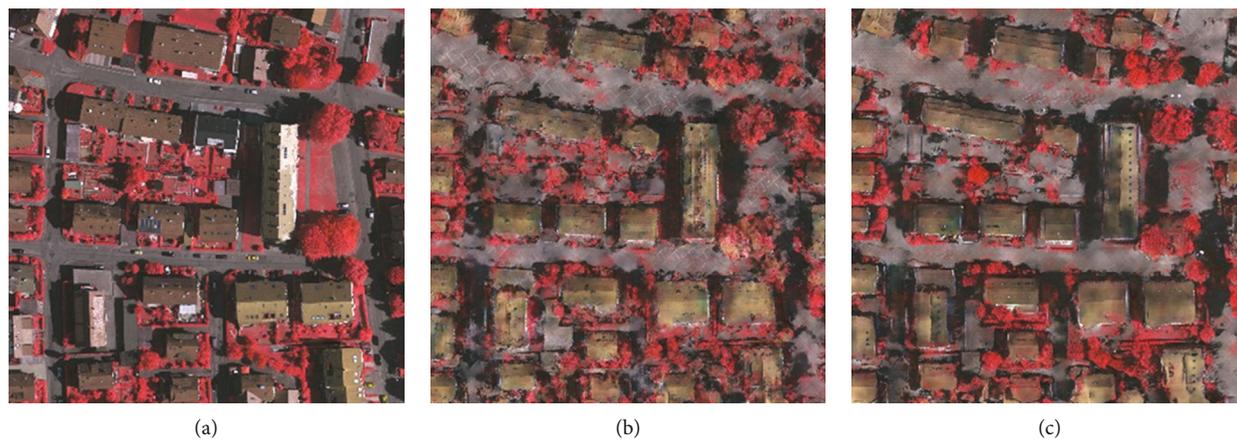


FIGURE 24: Generated true orthoimages with 256×256 tile size and 50% and 60% overlap: (a) real true orthoimage; (b) LiDAR DSM with 50% overlap; (c) LiDAR DSM with 60% overlap.

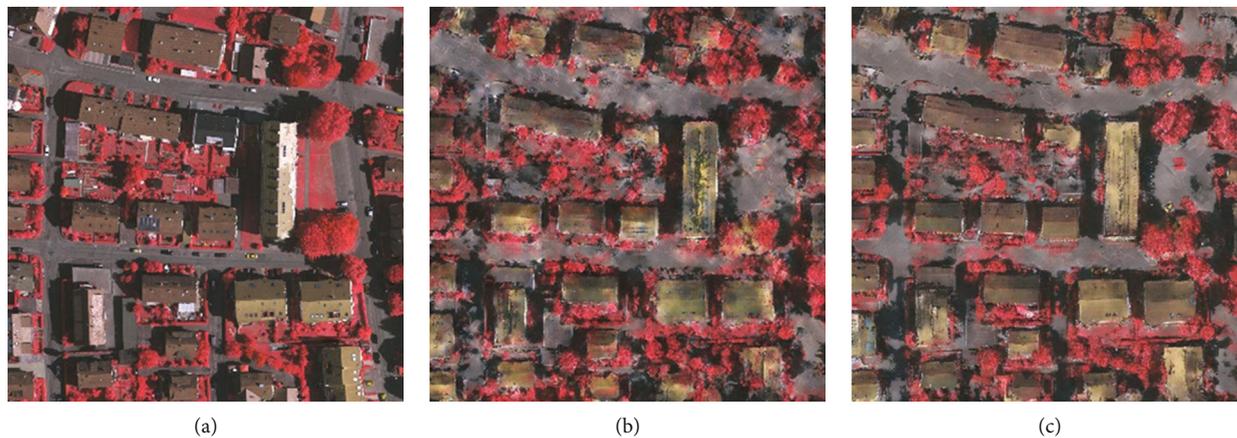


FIGURE 25: Generated true orthoimages with 512×512 tile size and 50% and 70% overlap: (a) real true orthoimage; (b) LiDAR DSM with 50% overlap; (c) LiDAR DSM with 70% overlap.

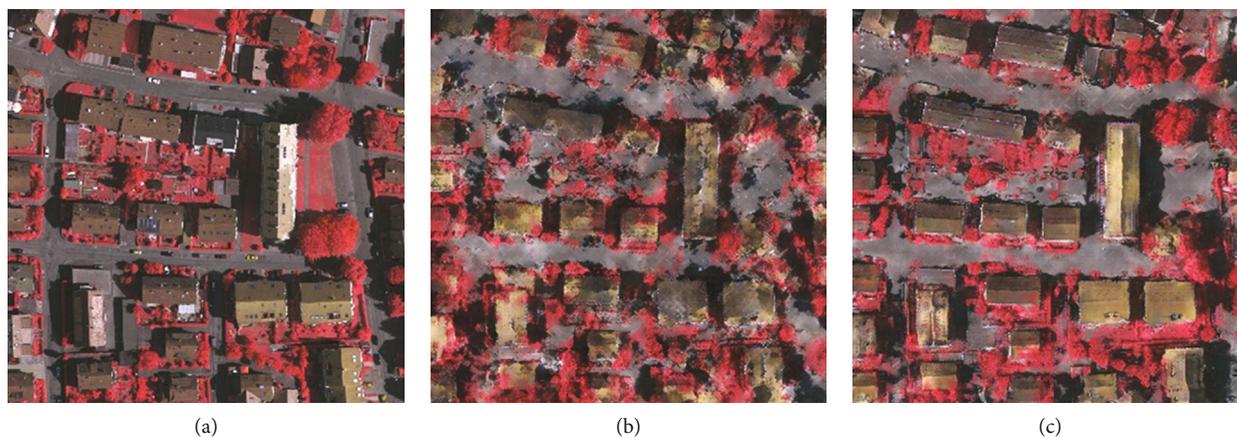


FIGURE 26: Generated true orthoimages with 1024×1024 tile size and 50% and 80% overlap: (a) real true orthoimage; (b) LiDAR DSM with 50% overlap; (c) LiDAR DSM with 80% overlap.

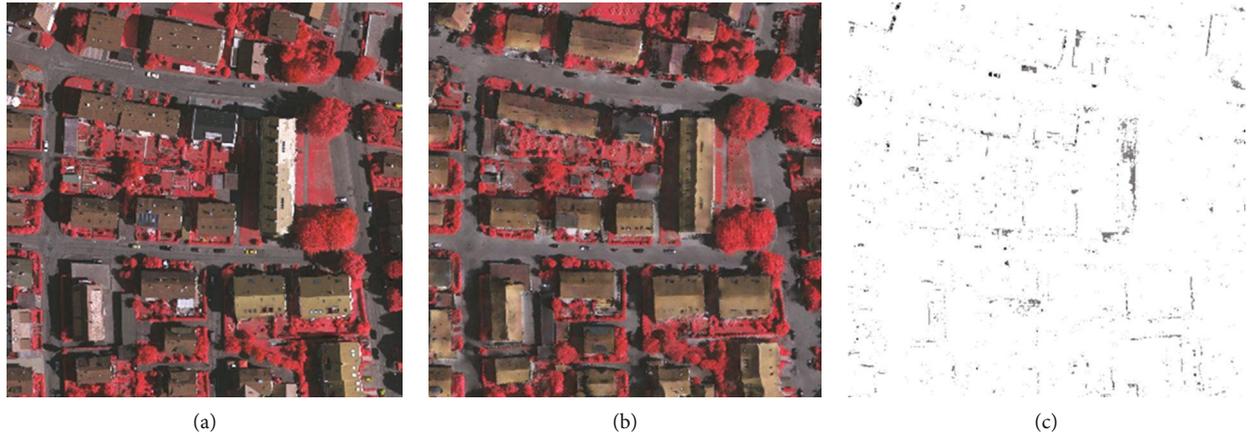


FIGURE 27: Generated true orthoimages with 256×256 tile size and 80% overlap: (a) real true orthoimage; (b) LiDAR intensity, LiDAR DSM, and B/W label; (c) difference image between (a) and (b).

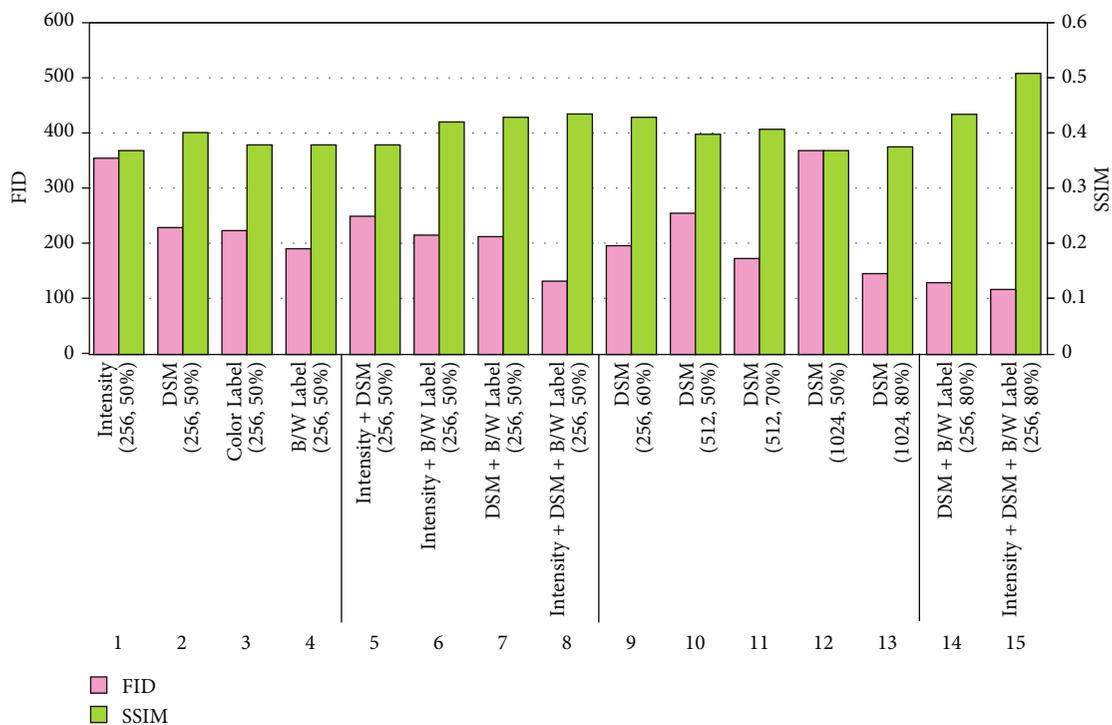


FIGURE 28: FID and SSIM of Vaihingen test datasets.

4. Conclusions and Discussions

Demand for true orthoimages is increasing since true orthoimages are important and useful products in the geospatial information system. However, generating true orthoimages is a challenging task using the conventional methods due to the complicated procedures involved in the process such as occlusion detection and recovery. This paper proposed a DL method for true orthoimage generation by utilizing the GAN-based Pix2Pix model. The crucial issues in DL are to improve the training efficiency and performance. The important factors in DL are characteristics, quantity, and quality of the training data. In this regard, the LiDAR data is one of the

most essential data because LiDAR data provides precisely georeferenced 3D coordinates and intensity information based on the orthogonal projection. Therefore, both geometric and physical information could be utilized to improve the performance of the DL model training especially for generating true orthoimages. Intensive and systematic experiments were performed to find the optimal condition of the training datasets to generate true orthoimages including data partition size, overlap, and integration of multimodal data (i.e., LiDAR intensity, DSM, and label).

In order to train the CNN-based DL models such as GAN, the LiDAR data is to be resampled by interpolation. However, the boundaries of the building (i.e., break lines or

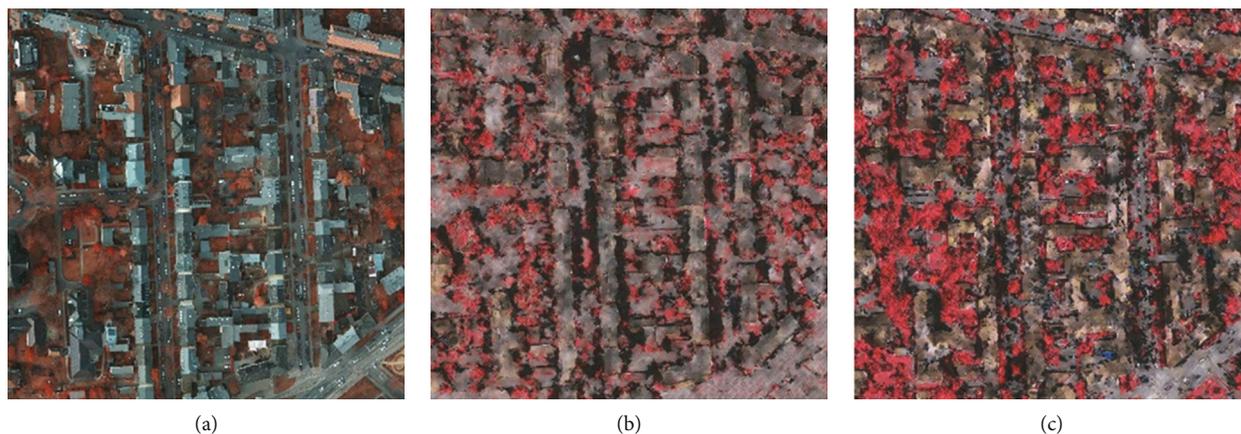


FIGURE 29: Generated true orthoimages with 256×256 tile size and 50% overlap: (a) real true orthoimage; (b) image matching DSM; (c) color label.

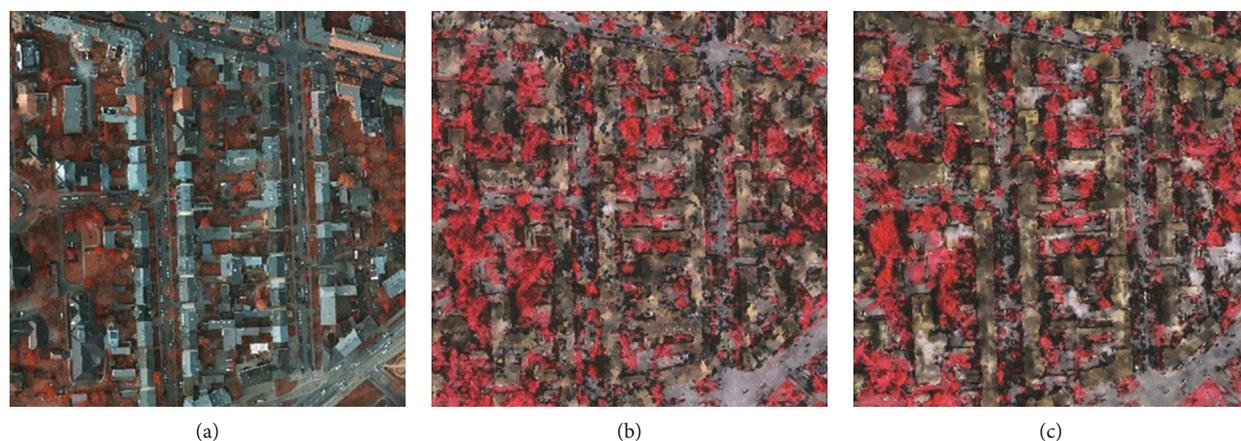


FIGURE 30: Generated true orthoimages with 256×256 tile size and 50% overlap: (a) real true orthoimage; (b) B/W label; (c) image matching DSM and B/W label.

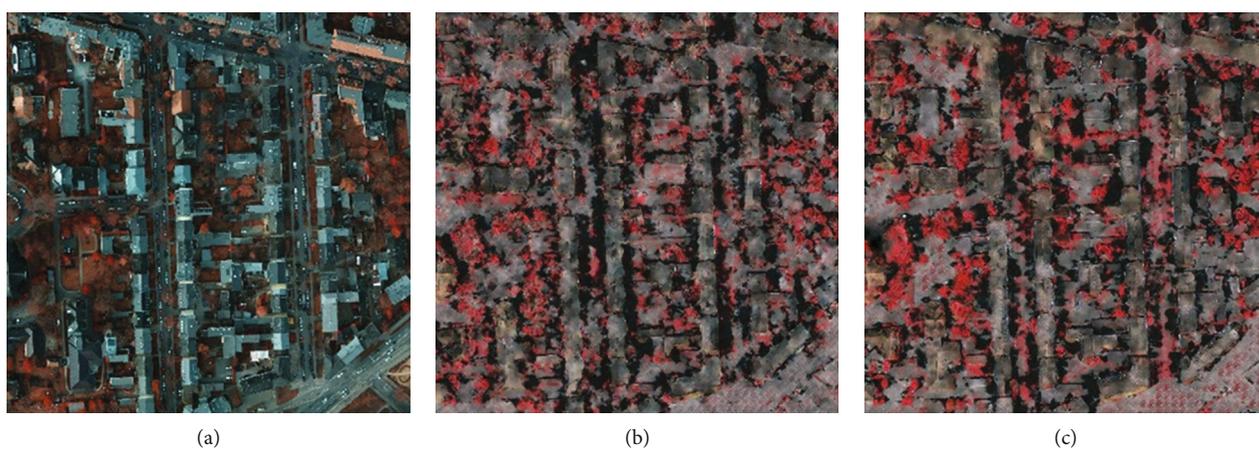


FIGURE 31: Generated true orthoimages with 256×256 tile size and 50% and 60% overlap: (a) real true orthoimage; (b) image matching DSM with 50% overlap; (c) image matching DSM with 60% overlap.

form lines) are not preserved due to the interpolation. In this regard, we carried out several experiments with different combinations of the training datasets to examine the contri-

bution of each individual data. We concluded that the label contributes to preserving the boundaries of the objects. Therefore, combination of the multimodel datasets having

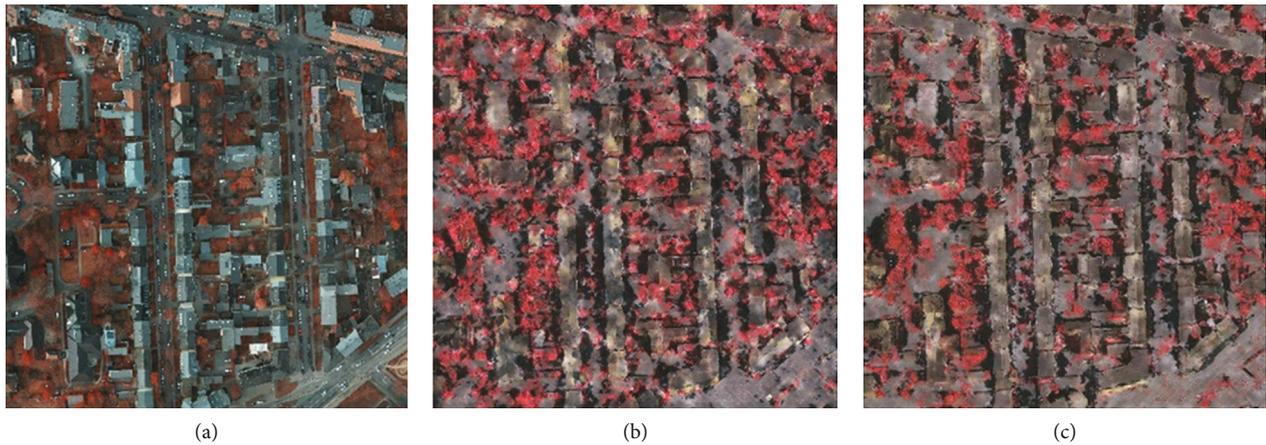


FIGURE 32: Generated true orthoimages with 512×512 tile size and 50% and 70% overlap: (a) real true orthoimage; (b) image matching DSM with 50% overlap; (c) image matching DSM with 70% overlap.

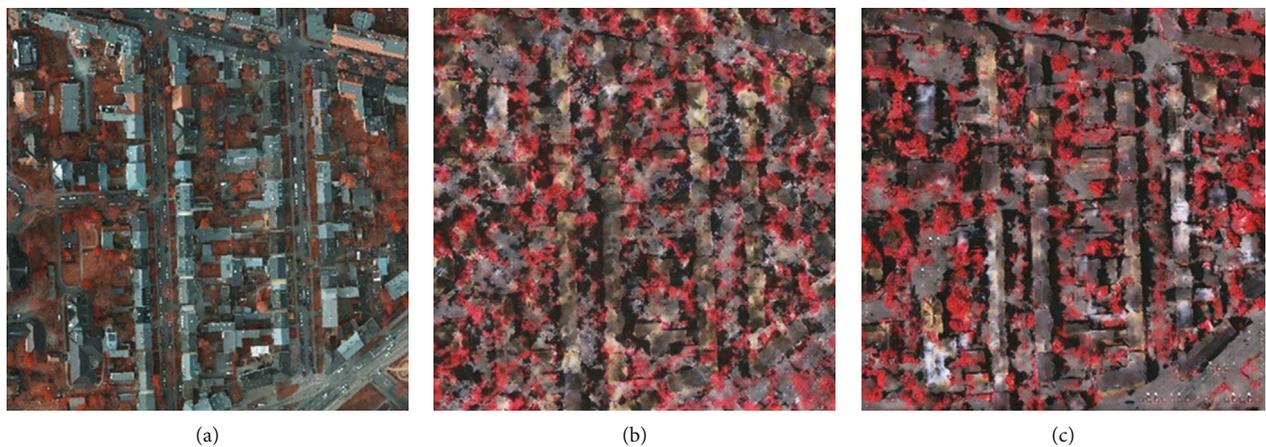


FIGURE 33: Generated true orthoimages with 1024×1024 tile size and 50% and 80% overlap: (a) real true orthoimage; (b) image matching DSM with 50% overlap; (c) image matching DSM with 80% overlap.

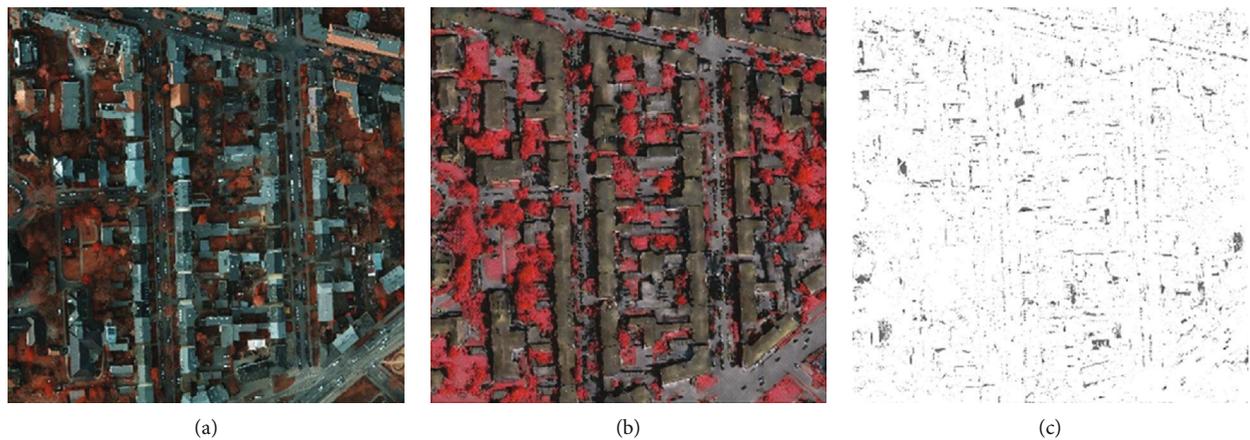


FIGURE 34: Generated true orthoimages with 256×256 tile size and 80% overlap: (a) real true orthoimage; (b) image matching DSM and B/W label; (c) difference image between (a) and (b).

different characteristics could result in mutual compensation of each individual data. In consequence, the synergistic effect during training DL models could be achieved.

Human beings have the ability to integrate various information to improve the learning process and the training performance on numerous applications. The same concept

TABLE 3: Evaluation of training performance for Potsdam unseen datasets.

| Experiment | Training data type | Tile size (pixel) | Overlap (%) | FID | SSIM |
|------------|--------------------|-------------------|-------------|--------|-------|
| 1 | DSM | 256 × 256 | 50 | 451.39 | 0.375 |
| 2 | Color label | 256 × 256 | 50 | 510.94 | 0.352 |
| 3 | B/W label | 256 × 256 | 50 | 515.57 | 0.363 |
| 4 | DSM+B/W label | 256 × 256 | 50 | 519.37 | 0.384 |
| 5 | DSM | 256 × 256 | 60 | 515.69 | 0.377 |
| 6 | DSM | 512 × 512 | 50 | 498.72 | 0.368 |
| 7 | DSM | 512 × 512 | 70 | 442.75 | 0.377 |
| 8 | DSM | 1024 × 1024 | 50 | 526.68 | 0.346 |
| 9 | DSM | 1024 × 1024 | 80 | 492.61 | 0.370 |
| 10 | DSM+B/W label | 256 × 256 | 80 | 373.03 | 0.387 |

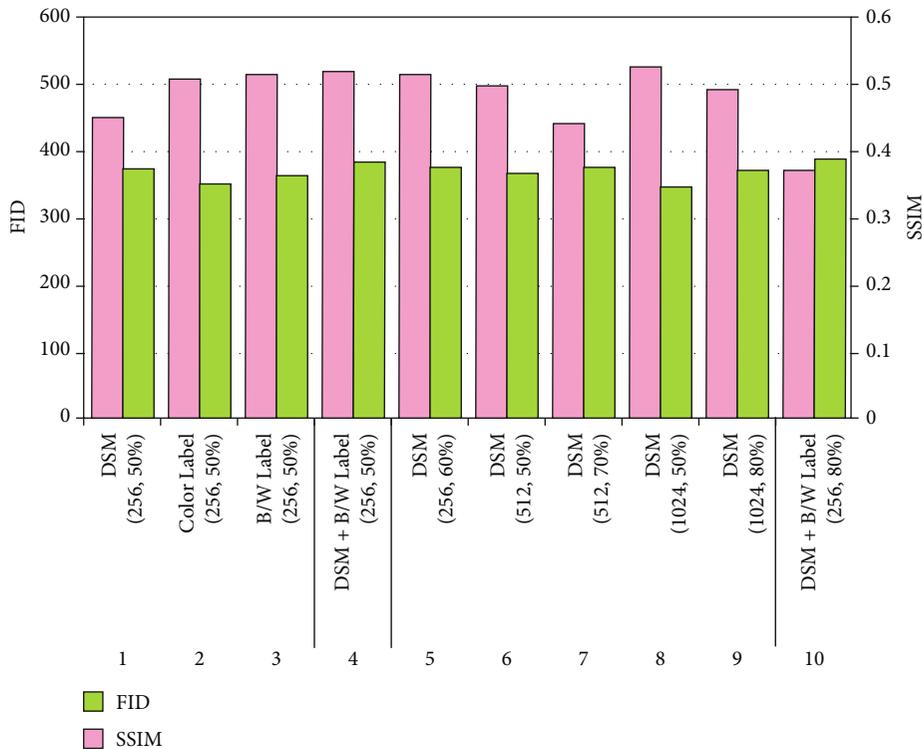


FIGURE 35: FID and SSIM of Potsdam unseen datasets.

could be applicable to DL. The major problem of the DL is data dependency. In other words, the results generated from the same DL model would be different even if the same training data is used. In this matter, various experiments were performed to provide trained models with different settings of the training datasets which could be provided including different tile sizes with overlap rates and different combinations of the datasets. It is obvious that the partitioning with overlap of the training datasets is beneficial because increasing the number of the datasets and diversity is desirable for training DL models. Also, training with multimodal datasets could provide improved results due to synergistic effect of each data. However, determining the optimal condition is not a

trivial task. Based on the experiments with the specific dataset used in this paper, the case of using smaller tile size with higher overlap rate and combining all data results in the best training performance.

In most cases, labels are essential in DL. Sometimes, the labels are not available and intensive work such as manual screen digitizing of the orthoimages is required to obtain the label information. However, the labels could be created efficiently using the digital vector maps as shown in Figure 38. Since the digital maps are composed of classified layers, various labels could be created by selecting appropriate layers for specific purposes. Different colors of the features can be assigned to the label.

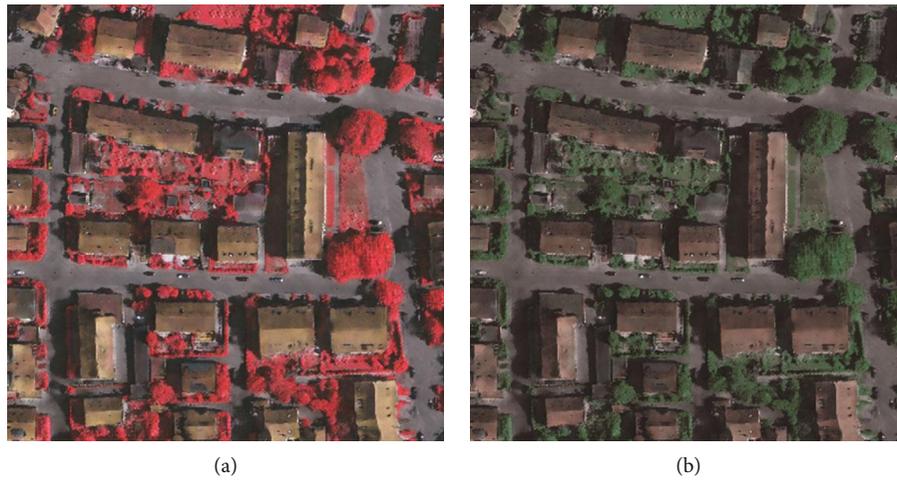


FIGURE 36: IR to natural color conversion of generated true orthoimage of test data: (a) IR true orthoimage; (b) pseudonatural color.

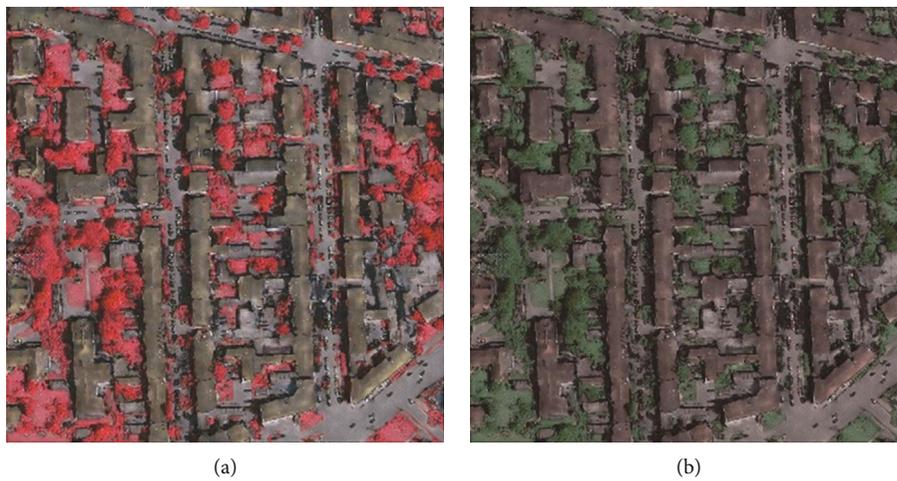


FIGURE 37: IR to natural color conversion of generated true orthoimage of unseen data: (a) IR true orthoimage; (b) pseudonatural color.

TABLE 4: Initial and final weights of spectral conversion matrix.

| Spectral weights | Initial weights | Final weights after training | Learning rate | Epoch |
|--|---|---|---------------|-------|
| $\begin{bmatrix} w_{11} & w_{21} & w_{31} \\ w_{12} & w_{22} & w_{32} \\ w_{13} & w_{23} & w_{33} \end{bmatrix}$ | $\begin{bmatrix} 0.1 & 0.9 & 0.0 \\ 0.3 & 0.0 & 0.7 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$ | $\begin{bmatrix} 0.078 & 0.887 & -0.013 \\ 0.239 & -0.041 & 0.659 \\ 0.040 & 0.059 & 0.759 \end{bmatrix}$ | 0.1 | 100 |



FIGURE 38: Example of label created from digital vector map: (a) building, road, and vegetation layer of digital map; (b) label.

The ultimate goal of DL is that the trained models are to be applicable to other datasets (i.e., unseen or new datasets) that are not involved in the model training. Unfortunately, use of the trained or pretrained models might be limited to the specific datasets that have similar properties to the training datasets. Therefore, various trained models to generate true orthoimages with multimodal datasets that take into account the regional and topographic properties could be more useful. In this matter, this paper might contribute to some strategies to improve the performance of the DL by providing an appropriate form of the training datasets.

Data Availability

The datasets generated during the current study are available from the corresponding author on reasonable request. The Vaihingen and Potsdam datasets are provided by the ISPRS Test Project on Urban Classification, 3D Building Reconstruction, and Semantic Labeling, <https://www2.isprs.org/commissions/comm2/wg4/benchmark>.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2018R1D1A1B07048732). The Vaihingen dataset was provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF) [23]: <http://www.ifp.uni-stuttgart.de/dgpf/DKEPAllg.html>.

References

- [1] E. Mikhail, J. Bethel, and J. McGlone, *Introduction to Modern Photogrammetry*, John Wiley & Sons, 2001.
- [2] K. Novak, "Rectification of digital imagery," *Photogrammetric Engineering and Remote Sensing*, vol. 58, no. 3, pp. 339–344, 1992.
- [3] G. Zhou, W. Schickler, A. Thorpe, P. Song, W. Chen, and C. Song, "True orthoimage generation in urban areas with very tall buildings," *International Journal of Remote Sensing*, vol. 25, no. 22, pp. 5163–5180, 2004.
- [4] Y. Liu, X. Zheng, G. Ai, Y. Zhang, and Y. Zue, "Generating a high-precision true digital orthophoto map based on UAV images," *ISPRS International Journal of Geo-Information*, vol. 7, no. 9, pp. 1–15, 2018.
- [5] M. Haggag, M. Zahran, and M. Salah, "Towards automated generation of true orthoimages for urban areas," *American Journal of Geographic Information System*, vol. 7, no. 2, pp. 67–74, 2018.
- [6] F. Amhar, J. Jansa, and C. Ries, "The generation of true orthophotos using a 3D building model in conjunction with a conventional DTM," *International Archives of Photogrammetry and Remote Sensing*, vol. 32, Part 4, pp. 16–22, 1988.
- [7] J. Rau, N. Chen, and L. Chen, "True orthophoto generation of built-up areas using multi-view images," *Photogrammetric Engineering and Remote Sensing*, vol. 68, no. 6, pp. 581–588, 2002.
- [8] K. Bang and C. Kim, "A new true ortho-photo generation algorithm for high resolution satellite imagery," *Korean Journal of Remote Sensing*, vol. 26, no. 3, pp. 347–359, 2010.
- [9] A. Habib, E. Kim, and C. Kim, "New methodologies for true orthophoto generation," *Photogrammetric Engineering and Remote Sensing*, vol. 73, no. 1, pp. 25–36, 2007.
- [10] P. Kuzmin, A. Korytnik, and O. Long, "Polygon-based true orthophoto generation," in *International Society for Photogrammetry and Remote Sensing (ISPRS) Congress*, pp. 529–531, Istanbul, Turkey, 2004.
- [11] H. Gharibi and A. Habib, "True orthophoto generation from aerial frame images and LiDAR data: an update," *Remote Sensing*, vol. 10, no. 4, p. 581, 2018.
- [12] E. Yoo and D. C. Lee, "True orthoimage generation by mutual recovery of occlusion areas," *GIScience & Remote Sensing*, vol. 53, no. 2, pp. 227–246, 2016.
- [13] G. Zhou, W. Chen, J. Kelmelis, and D. Zhang, "A comprehensive study on urban true orthorectification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 9, pp. 2138–2147, 2005.
- [14] A. Gunay, H. Arefi, and M. Hahn, "Semi-automatic true orthophoto production by using LIDAR data," in *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 2873–2876, Barcelona, Spain, 2007.
- [15] Y. Sheng, "Minimising algorithm-induced artefacts in true ortho-image generation: a direct method implemented in the vector domain," *The Photogrammetric Record*, vol. 22, no. 118, pp. 151–163, 2007.
- [16] L. Barazzetti, M. Brovelli, and M. Scaioni, "Generation of true-orthophotos with LiDAR high resolution digital surface models," *The Photogrammetric Journal of Finland*, vol. 21, no. 1, pp. 26–36, 2008.
- [17] H. Oliveira, M. Galo, and A. Dal Poz, "Height-gradient-based method for occlusion detection in true orthophoto generation," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 11, pp. 2222–2226, 2015.
- [18] C. Zhong, H. Li, and X. Huang, "A fast and effective approach to generate true orthophoto in built-up area," *Sensor Review*, vol. 31, no. 4, pp. 341–348, 2011.
- [19] H. Oliveira, A. Poz, M. Galo, and A. F. Habib, "Surface gradient approach for occlusion detection based on triangulated irregular network for true orthophoto generation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 2, pp. 443–457, 2018.
- [20] H. Bagheri and S. Sadeghian, "Ortho image and DTM generation with intelligent methods," in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-1/W3, pp. 475–480, Tehran, Iran, 2013.
- [21] Y. Shin, S. Hyung, and D. C. Lee, "True orthoimage generation from LiDAR intensity using deep learning," *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, vol. 38, no. 4, pp. 363–373, 2020.
- [22] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial nets," in *Proceedings of the International Conference on Advances in Neural Information Processing Systems (NIPS)*, pp. 2672–2680, Montreal, Canada, 2014.

- [23] M. Cramer, "The DGPF-test on digital airborne camera evaluation overview and test design," *Photogrammetrie - Fernerkundung - Geoinformation*, vol. 2010, no. 2, pp. 73–82, 2010.
- [24] F. Rottensteiner, G. Sohn, J. Jung et al., "The ISPRS benchmark on urban object classification and 3D building reconstruction," *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. I-3, pp. 293–298, 2012.
- [25] D. Lee, Y. Shin, and D. C. Lee, "Land cover classification using SegNet with slope, aspect, and multidirectional shaded relief images derived from digital surface model," *Journal of Sensors*, vol. 2020, Article ID 8825509, 21 pages, 2020.
- [26] P. Isola, J. Zhu, T. Zhou, and A. Efros, "Image-to-image translation with conditional adversarial networks," in *2017 IEEE conference on computer vision and pattern recognition (CVPR)*, pp. 5967–5976, Honolulu, USA, 2017.
- [27] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: an overview," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 53–65, 2018.
- [28] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015), Lecture Notes in Computer Science*, vol. 9351, Springer, 2015.
- [29] P. Nagorny, T. Lacombe, H. Favrelière et al., "Generative adversarial networks for geometric surfaces prediction in injection molding: performance analysis with discrete modal decomposition," in *2018 IEEE International Conference on Industrial Technology (ICIT)*, pp. 1514–1519, Lyon, France, 2018.
- [30] V. Nair and G. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML)*, pp. 807–814, Haifa, Israel, 2010.
- [31] A. Maas, A. Hannun, and A. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proceedings of the 30th International Conference on Machine Learning (ICML)*, Atlanta, USA, 2013.
- [32] C. Aguilera, X. Soria, A. D. Sappa, and R. Toledo, "RGBN multispectral images: a novel color restoration approach," in *International Conference on Practical Applications of Agents and Multi-Agent Systems*, pp. 1–9, Porto, Portugal, 2017.
- [33] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [34] D. Brunet, E. Vrscay, and Z. Wang, "On the mathematical properties of the structural similarity index," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1488–1499, 2012.
- [35] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local nash equilibrium," in *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS)*, pp. 6629–6640, Long Beach, CA, USA, 2017.
- [36] A. Borji, "Pros and cons of GAN evaluation measures," *Computer Vision and Image Understanding*, vol. 179, pp. 41–65, 2019.